Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase

by

Andrew J. Serban

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2018 by the
Graduate Supervisory Committee:

Rebekka Wachter, Chair
Marcia Levitus, Co-Chair
Kevin Redding
Wade Van Horn

ARIZONA STATE UNIVERSITY

August 2018

ABSTRACT

Rubisco activase (Rca) from higher plants is a stromal ATPase essential for reactivating Rubiscos rendered catalytically inactive by endogenous inhibitors. Rca's functional state is thought to consist of ring-like hexameric assemblies, similar to other members of the AAA+ protein superfamily. However, unlike other members, it does not form obligate hexamers and is quite polydisperse in solution, making elucidation of its self-association pathway challenging. This polydispersity also makes interpretation of traditional biochemical approaches difficult, prompting use of a fluorescence-based technique (Fluorescence Correlation Spectroscopy) to investigate the relationship between quaternary structure and function. Like cotton β Rca, tobacco β Rca appears to assemble in a step-wise and nucleotide-dependent manner. Incubation in varying nucleotides appears to alter the equilibrium between varying oligomers, either promoting or minimizing the formation of larger oligomers. High concentrations of ADP seem to favor continuous assembly towards larger oligomers, while assembly in the presence of ATP-γS (an ATP analog) appears to halt continuous assembly in favor of hexameric species. In contrast, assembly in the "Active ATP Turnover" condition (a mixture of ATP and ADP) appears to favor an almost equal distribution of tetramer and hexamer, which when compared with ATPase activity, shows great alignment with maximum activity in the low μM range. Despite this alignment, the decrease in ATPase activity does not follow any particular oligomer, but rather decreases with increasing aggregation, suggesting that assembly dynamics may regulate ATPase activity, rather than the formation/disappearance of one specific oligomer. Work presented here also indicates that all oligomers larger than hexamers are catalytically inactive, thus providing support for the idea that they may serve

as a storage mechanism to minimize wasteful hydrolysis. These findings are also supported by assembly work carried out on an Assembly Mutant (R294V), known for favoring formation of closed-ring hexamers. Similar assembly studies were carried out on spinach Rca, however, due to its aggregation propensity, FCS results were more difficult to interpret. Based on these findings, one could argue that assembly dynamics are essential for Rca function, both in ATPase and in regulation of Rubisco carboxylation activity, thus providing a rational for Rca's high degree of polydispersity.

DEDICATION

To my lovely wife Cristina, who supported me through the peaks and the valleys of my graduate career, and to my two children, Jason and Emily, who have turned this career into an adventure.

ACKNOWLEDGMENTS

For my wife Cristina, who has shown me her never-ending support in pursuing my doctoral degree, even in difficult times. For my parents, Nelu and Lidia, and my brother Paul, who have constantly encouraged me to put my best foot forward in all academic pursuits. For my lab members, both past and present, in both the Wachter and Levitus labs who sat through many of my group meeting talks and provided key insight and suggestions on how to push the research forward. For Dr. Nathan J. Henderson, who helped me get my footing at the start of my graduate career by answering far too many of my questions. For Dr. Agnieszka Kuriata who trained me on the biochemical portion of the project and for Dr. Manas Chakraborty who trained me on the FCS instrument. For my colleagues, Eduardo Espiritu and Dr. Bryan Donaphon, who always had an encouraging word to say when experiments failed. For my advisors, Rebekka and Marcia, who gave me the opportunity to do research in their lab and who have helped me grow into the scientist I am today.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

Photosynthesis is considered by some to be one of the most important process on Earth today, due in part to its role in the biofuel and energy sectors, but primarily due to its importance in producing the amount of food necessary for sustaining our ever-growing population [1]. Given this increase in global food demand, much emphasis as of late has been placed on better understanding the mechanics of the photosynthesis process, a trend that was not always the case. In years past, much of the focus was placed on better understanding the external factors governing plant growth. One such factor that received much attention during, what experts refer to as, the "green revolution" was fertilizers [2]. Experts argued that having a complete understanding of all micro-nutrients required for effective plant growth would lead to better designed fertilizers, which in turn would promote greater crop yields [2,3]. Time has shown that they were right in this regard. Another focus during this revolution was placed on characterizing and understanding those diseases most affecting food crops and designing chemical agents to minimize their effects. Like fertilizers, targeting and preventing disease also drastically improved crop yields [4]. However, the greatest increase in crop yields during the "green revolution" was observed when researchers focused their work on increasing the genetic yield potential. This term refers to having a thorough understanding of all optimal management practices (total incident solar irradiation, light interception efficiency, conversion efficiency, and partitioning efficiency) for the particular crop plant such that one can avoid subjecting said crop to both biotic and abiotic stresses [2,4].

Due to the discoveries and improvements made during the revolution, food crop production has been able to stay in step with increasing demands from an increasing population. However, as of late, this increasing trend in crop yield has grown stagnant, as made evident by a study looking at the average overall yields of rice plants over the span

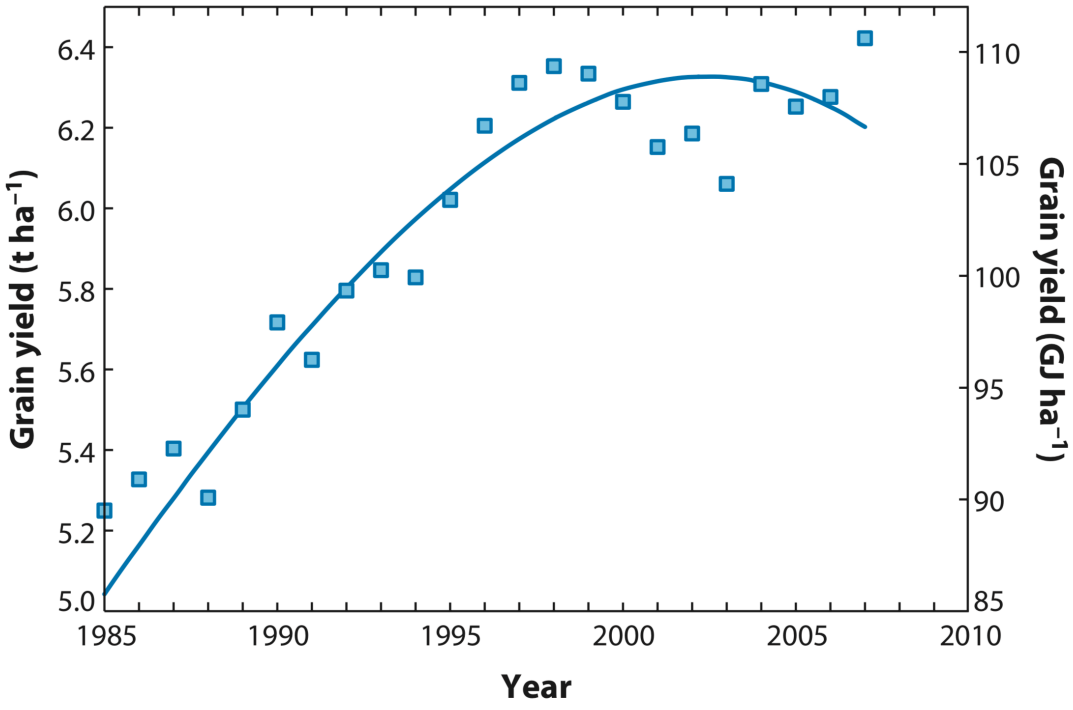


Figure 1.1: Annual average rice crop yields in China. The annual yield of rice crop plants per unit land area in China with each point being the average of all harvests. This figure is adapted from [2].

of 25 years [2,5]. From years 1987 to 1997, the average rice crop yields in China have increased from 5.4 t ha^{-1} to 6.4 t ha^{-1}, however, from 1997 to 2007, the yield has remained relatively constant, with small fluctuation from year to year (**Fig. 1.1**). This trend, which is seen for other food crops as well, suggests that optimization of those areas focused on in the "green revolution" have reached a point of diminishing returns and thus lack the potential for increasing further [2]. This does not bode well for our food producers considering the increasing demands on grains from countries like India and China [6], the

supply of which, the Food and Agriculture Organization of the United Nations predicts, needs to double by 2050 [2]. Therefore, researchers are arguing that meeting this demand will require a shift in focus towards cultivating a more complete understanding of the entire photosynthetic process at a molecular level and engineering aspects of it to be more efficient [7].

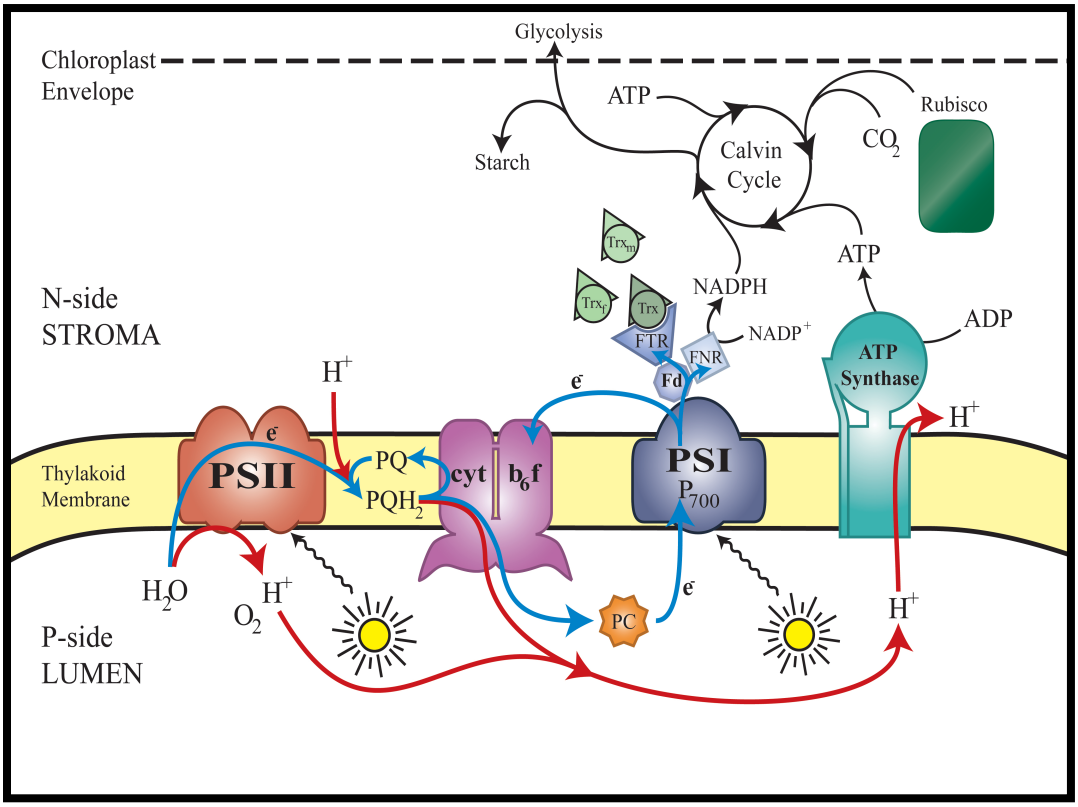


Figure 1.2: Electron transport chain found inside the chloroplast. Cartoon representation of the electron transport chain located inside the chloroplast of higher plants, courtesy of Dr. Watkins. PSII utilizes incoming light to shuttle electrons down the chain using PQ, generating PQH_2, which is further shuttled by the Cyt b_6f, to the PC, which acts as a reducing agent for PSI. PSI shuttles electrons up one branch of the electron transfer chain, up to F_x, which is then reduced by Fd, which is used by FNR to generate NADPH. Proton pumping occurs across the entire chain allowing for ATP production by ATP Synthase.

When discussions regarding re-engineering of photosynthesis arise, one of the key areas most researchers direct their attention to is the electron transport chain (**Fig. 1.2**).

This chain, found within the thylakoid membranes in the chloroplast, consists of many proteins whose ultimate role is to convert incoming solar radiation into useable chemical energy [8,9]. The first protein responsible for this conversion is photosystem II (PSII), a heterodimer with 68 chlorophyll antenna molecules and many more co-factors [8,10]. In brief, photosystem II utilizes its chlorophyll antenna network to capture an incoming photon and shuttles the resulting displaced electron down to what is known as it's special pair, P680, forming oxidized $P680^+$. The P680 nomenclature comes from the absorbance maximum of 680 nm identified for this pair [8]. Structurally, this special pair consists of two parallel chlorophyll molecules, forming a stacked dimer, which allows the continued shuttling of the electron up the electron transfer chain of the reaction center to a plastoquinone as the terminal acceptor [10,11]. Reduction of $P680^+$ is carried out by the oxygen evolving complex (OEC) via a water splitting mechanism resulting in formation of H^+ and O_2, which in turn allows the cycle to begin again [12,13]. This act of water splitting, allows for an active pumping of H^+ ions into the luminal space of the thylakoid. Once two electrons are shuttled through the cycle, plastoquinone is reduced to plastoquinol (PQ) and transferred into the PQ pool. From this pool, PQ is shuttled to the next protein in the chain, cytochrome b_6f, which is a dimeric protein tasked with facilitating the reduction of plastocyanin (PC) (**Fig. 1.2**) [11,14]. Once PQ has reduced PC, the resulting plastoquinone is shuttled back to the PSII complex to begin the process again. As with PSII, this reduction of PC by the b_6f complex results in active pumping of even more H^+ ions into the luminal space (**Fig. 1.2**) [14,15].

Like PSII, photosystem I (PSI), the next member of the chain, also uses its chlorophyll antenna network to capture incoming photons and shuttle the resulting electron

to its special pair. However, unlike PSII, PSI's special pair is termed P700, meaning that its absorbance maximum is at 700 nm instead of 680 nm [16]. Additionally, the reduction of $P700^+$ is carried out by the PC instead of the OEC, and the terminal acceptor for the electron going up the electron transfer chain of the reaction center is an 4Fe-4S cluster, termed F_x, instead of plastoquinone [16,17]. This change in terminal acceptor allows for reduction of ferredoxin, an FeS protein capable of quickly shuttling and housing electrons to and from different proteins [16]. In the case of the PSI, ferredoxin is used in conjunction with oxidized nicotinamide adenine dinucleotide phosphate ($NADP^+$) by ferredoxin-$NADP^+$ reductase (FNR) to generate oxidized ferredoxin and NADPH, allowing ferredoxin to be reused by PSI (**Fig. 1.2**) [11,16,17].

The last protein associated with electron transport chain is the ATP synthase, a complex which utilizes the proton motive forces (pmf) generated by the active pumping of H^+ molecules into the lumen to generate adenosine triphosphate (ATP) (**Fig. 1.2**) [18]. The mechanism behind ATP generation, in brief, consists of a ratchet-like system, where individual protons are loaded into the rotating c-ring, in turn rotating the shaft [19], known as the γ complex, and cycling through 3 different states, (open, tight, and loose) in the crown of the F1 domain, where the open site allows for ADP and phosphate binding, the tight state involves ATP formation, and the loose state allows for ATP release [11,18]. The ATP generated from this cycle, along with the NADPH generated by FNR, are fed into the Calvin-Benson-Bassham (CBB) Cycle (**Fig. 1.2**).

The Calvin-Benson-Bassham Cycle

Often referred to as the light-independent reactions, the primary role of the Calvin-Benson-Bassham cycle is to generate necessary precursor molecules used for glucose

synthesis in plants by combining sugar phosphate molecules with atmospheric CO_2. The key enzyme tasked with facilitating this process is known as ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) [20]. Some estimates speculate that nearly 50% of the total soluble protein content inside a leaf is comprised of Rubisco making it not only important but also highly abundant [21]. Rubisco carries out this process by combining RuBP, a 5 carbon sugar, with the incoming CO_2 molecule, to form a 6 carbon sugar intermediate, which in turn is broken down into two 3 carbon molecules, called 3-phosphoglycerate (3PGA) [20,22]. This molecule is later processed into glyceraldehyde-3-phosphate (G3P), a molecule used by plants to generate glucose. However, because RuBP needs to be regenerated for the cycle to begin again, 3 complete cycles are needs to generate 1 extra G3P that can be used for glucose production [22,23]. This carbon fixation step of the reaction carried out by Rubisco is considered to be the bottleneck reaction of the Calvin-Benson-Bassham (CBB) cycle and yet, is responsible for all of the biomass accumulation observed in the world today [21,23].

Rubisco: Structure and Function

There are four distinct types of Rubisco, forms I, II, III, and IV, however, form I is far more prevalent than the others [21]. Structurally, form I Rubisco is a hexadecamer, consisting of eight large subunits, roughly 55 kDa each, and eight small subunits, roughly 14 kDa each, denoted commonly as L8S8, where L stands for large and S for small [21]. Forms II and IV, in contrast to form I, lack the small subunit, and thus form a large subunit dimer (L2). Form III, unlike the others, uses five L2 dimers to form a pentamer of dimers (L10) [21,24] (**Fig. 1.3**). The substrate binding sites for all forms of Rubisco are generated at the interface of two anti-parallel large subunits, meaning that there are two substrate-

binding sites for every 2 large subunits. These large subunit dimers are oriented such that the binding site is solvent accessible when in the open state, allowing access for RuBP binding [21]. In Rubisco's closed state, the loop 6 region sits on top the binding site and is responsible for granting and preventing access to the active site [21,25]. It does so by

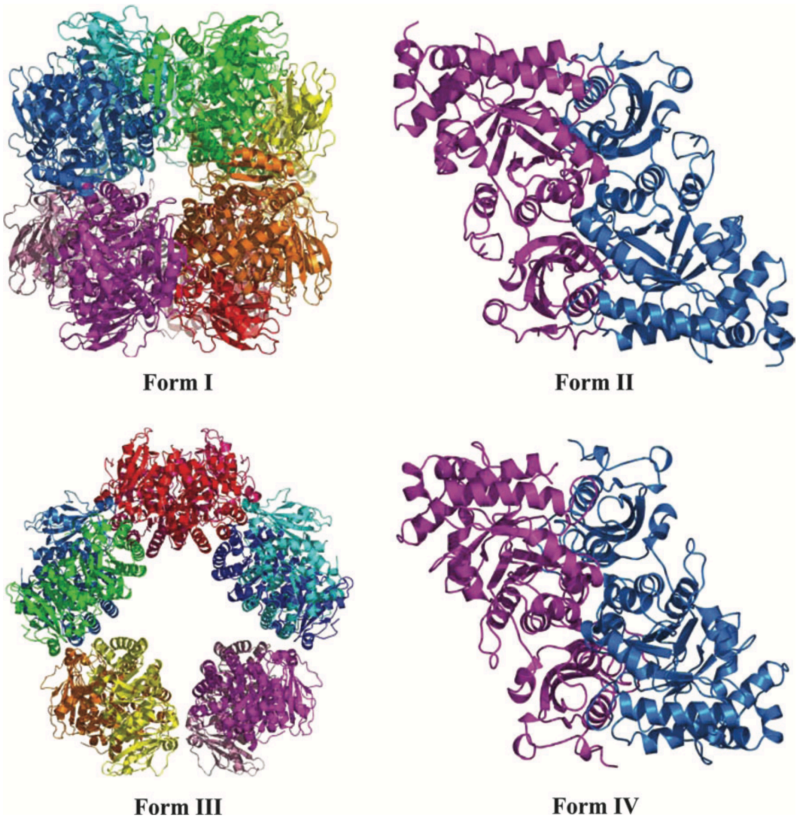


Figure 1.3: Various forms of Rubisco. Four different forms in the Rubisco family. Form I shows the traditional L8S8 conformation. Forms II and IV, which lack the small subunits, adopt the L2 conformation. Form III, unlike the rest, forms the L10 conformation, consisting of 5 L2 subunits. This figure is adapted from (24).

utilizing a specific lysine residue (Lys 334 spinach numbering) to form a hydrogen bond to the oxygens of the incoming CO_2 molecule, creating a hinge-like system (**Fig. 1.4**). Additionally, the entire large subunit pair undergoes slight structural rearrangements to help lock in the bound substrate, which in combination with the loop 6 region, generates the desired closed conformation [21,24].

In order to have proper substrate turnover, the active site must be in the open

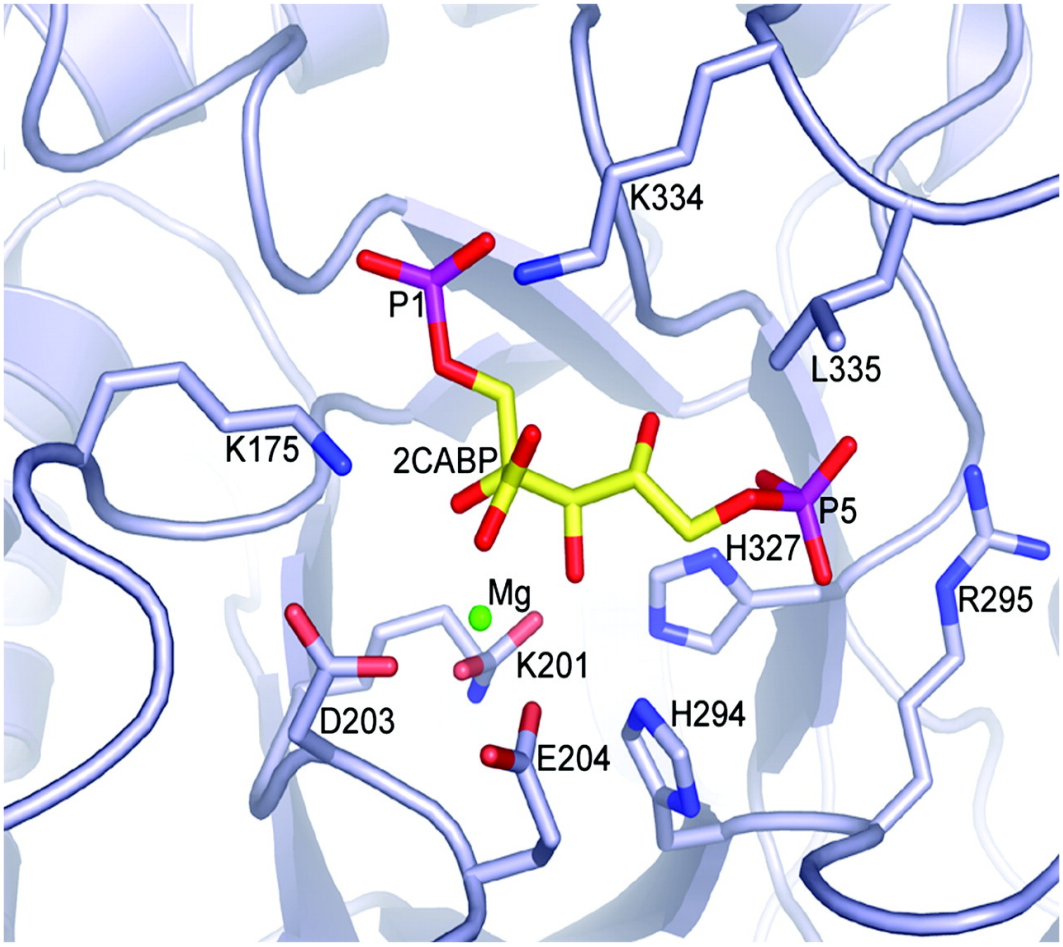


Figure 1.4: Understanding the active site of Rubisco. Identification of key residues involved in generating the active site of Rubisco. Carbamylation occurs at the K201 residue, which then coordinates Mg^{2+} to stabilize RuBP. Additionally, the K334 residue located on loop 6 plays an important role in substrate coordination. This figure is adapted from (32).

conformation and get carbamylated, meaning that a particular lysine residue (Lys 201 spinach numbering) in the active site must covalently bind CO_2 (a different CO_2 molecule than that being captured from the atmosphere) and have a Mg^{2+} ion in close proximity (**Fig. 1.4**) [25,26]. This carbamylation process transitions Rubisco from an "E" state, where E refers to enzyme only, into a state called ECM which means Enzyme CO_2 Magnesium (**Fig. 1.5**). The free Mg^{2+} ions help stabilize the carbamylated lysine as well as coordinate the

incoming CO_2 molecule that is going to be used in formation of 3PGA [21]. If carbamylation does not occur prior to RuBP binding, Rubisco will be inhibited and unable to turnover, populating a state referred to as ER, Enzyme RuBP (**Fig. 1.5**), which is believed to serve as a regulator of Rubisco activity [24,25].

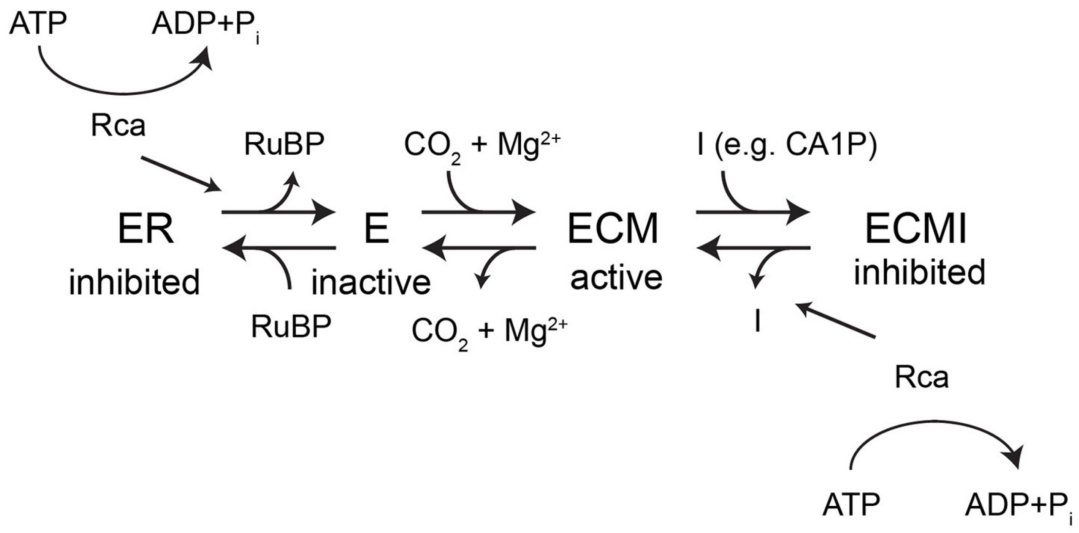


Figure 1.5: Individual steps in the Rubisco activation/inactivation process. Schematic showing the various steps related to different states of Rubisco over its life-cycle. In the inhibited form (ER), Rca is recruited to reactive Rubisco (E) via ATP hydrolysis, at which point carbamylation can occur resulting in active Rubisco (ECM). If an inhibitor were to bind to Rubisco (ECMI), Rca would be recruited again to reactivate Rubisco. This figure is adapted from (40).

Limitations and Inhibition:

Despite having such an important biological role, Rubisco has two major limitations. The first is Rubisco's propensity to incorporate O_2 into its binding site instead of the desired CO_2. In doing so, Rubisco generates 1 molecule of 3PGA and 1 molecule of 2-phosphoglycolate [21]. This latter product must be recycled by the plant via the photorespiration pathway and is considered to be quite energetically costly and wasteful [27,28]. In addition, Rubisco suffers from a secondary limitation in that it has a high

propensity to get inhibited, at times even by its own substrate. As mentioned above, there are a strict series of steps that must occur prior to RuBP binding, particularly the carbamylation of Lys201 and the presence of Mg^{2+}. If RuBP binds without these conditions being met, it becomes an inhibitor rather than a substrate. This is considered to be a substantial limitation because once in this inactive state, Rubisco must recruit the aid of an activator protein, termed, Rubisco activase (Rca) to reactivate Rubisco to its catalytic form [29]. However, despite being perceived as a limitation, many scientists claim that it might be a method of Rubisco regulation. For example, another inhibitor, 2-carboxy-D-arabinitol 1-phosphate (CA1P), is in high abundance when daylight grows faint, allowing the plant to suppress Rubisco activity [29,30,31]. Conversely, during high light, CA1P is removed from the active site by Rca and is subsequently dephosphorylated, reducing its binding propensity but still housing it nearby for immediate re-phosphorylation when light fades [29,31]. This same sort of regulation is thought to occur with other types of inhibitors, such as xylulose-1,5-bisphosphate (XuBP), which is generated by mis-protonation of the intermediate, or pentadiulose-1,5-bisphosphate (PDBP) which is produced downstream of 2-phosphoglycolate formation, however, unlike CA1P, their biological function is less understood [29,30,32,33]. Regardless of the inhibitor, it is important to remember that Rca has a critical role in removing said inhibitor.

Given these limitations, the past few decades have seen great efforts from researchers attempting to re-engineer Rubisco in order to make it more effective in its role [34,35]. In particular, researchers tried to design mutants capable of better distinguishing between CO_2 and O_2 as well as mutants which have higher turnover rates [36]. Despite their efforts, their findings suggested that making Rubisco more effective at differentiating

between the two molecules caused lower turnover rates, while conversely, mutants with higher turnover rates were seen to be less effective at distinguishing between CO_2 and O_2, causing them to conclude that the balance nature has reached in optimizing Rubisco is indeed the best of both traits [36,37]. It is worth noting that one of the biggest limitations in screening different mutant constructs for improved function was the difficulty in introducing the particular mutation into the protein sequence due to the large subunit of Form I Rubisco being encoded by a chloroplast gene. Given this issue, researchers made strong efforts to recombinantly express Rubisco in bacterial host but failed, due to their inability to co-express all necessary chaperone proteins necessary for properly assembly Rubisco [36,38,39]. Therefore, at the time, the focus shifted from re-engineering Rubisco only, to co-engineering Rubisco and Rca together, which required a thorough understanding of the newly discovered Rca protein. However, as of less than a year ago, this expression problem has been solved as a German group from the Max Planck Institute, opening the door once again to re-engineering efforts [38,39].

The Rubisco Activase Family

There are currently three distinct classes of activases, green-type, red-type, and the CbbQO-type, but given that green-type activases were discovered first in the late 80's and are most abundant, a greater emphasis was placed on characterizing them, both structurally and functionally [40,41]. Despite all the knowledge gained thus far, the story on activases is far from complete. Based on their sequences, all three Rca types belongs to the AAA+ (ATPases Associated with various cellular Activities) protein super family [40]. Structurally, all three Rca contains an AAA+ module, however, differ more in their N and C terminal extensions [25,40]. The AAA+ module is comprised of an α/β subdomain, seen

in teal, and an α-helical subdomain, seen in yellow [42] (**Fig. 1.6**). As is the case with most

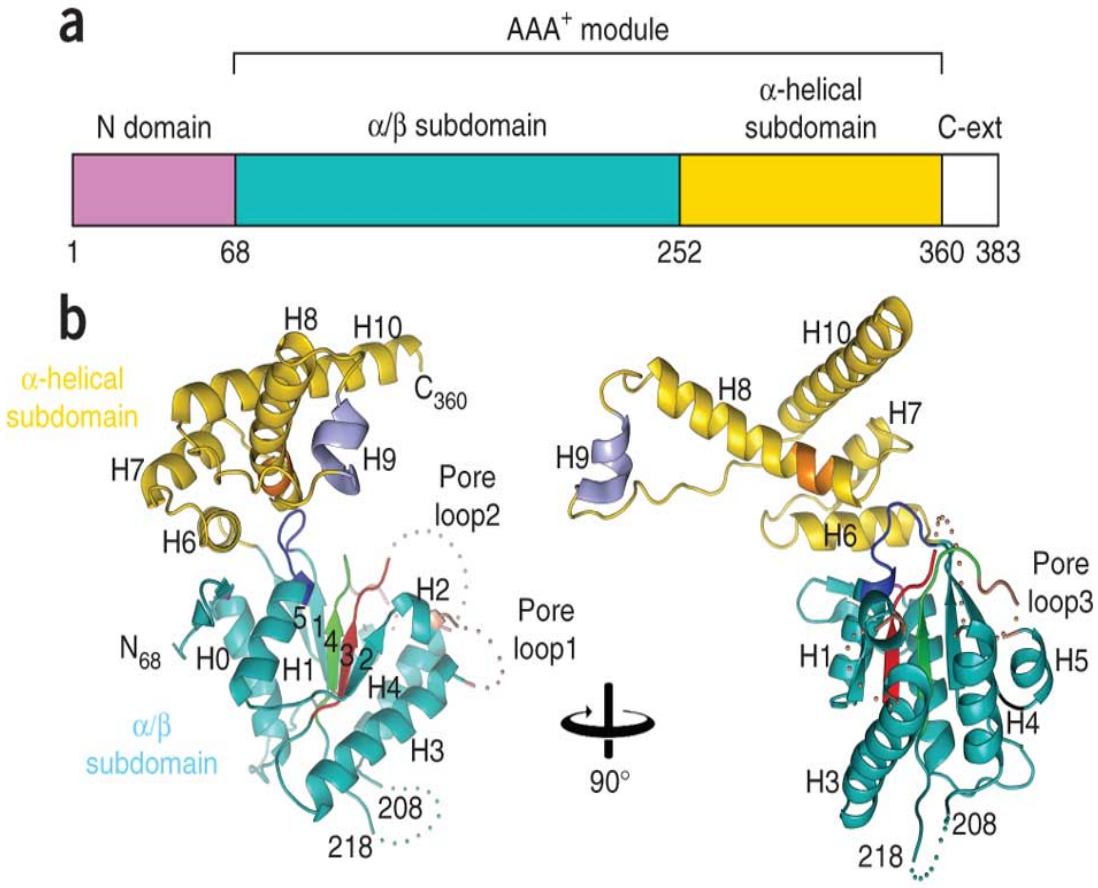


Figure 1.6: Domain organization & structural characterization of green-type activase. A) the domain organization in Rca showing the canonical AAA+ module, with the two respective subdomains shown in teal and yellow, as well as the N-terminal (pink) and C-terminal (white) extensions. These domains are color coded with **b**) the monomeric structure of Rca, highlighting the Walker A (dark blue) and Walker B (red) motifs as well as Sensor I (green) and Sensor II (orange). This figure is adapted from (42).

AAA+ proteins, including activases, the binding site for the nucleotide is located in between the two subdomains of one protomer and is further shielded by the interface of the adjacent subunit [43]. This interface interaction between adjacent subunits is quite important because there are a series of important residue interactions that help form and stabilize the binding site.

Within the AAA+ domain of activases, there are iconic motifs seen in most other AAA+ proteins, specifically, the Walker A and B motif along with Sensors 1 and 2 and the Arginine finger, all of which help stabilize the bound nucleotide [25]. In brief, the Walker A motif, seen in dark blue (**Fig. 1.6**), plays an integral part in properly coordinating the nucleotide in the binding site by its interaction with the phosphates [25]. The particular sequence of this motif consists of GXXXXGK, where K is most crucial [43,44]. Similarly, the Walker B motif, seen in red, also plays a critical role in nucleotide coordination, but unlike the Walker A motif, it also plays a role in hydrolysis. In most AAA+ protein, including red-type and CbbQO, the canonical sequence is hhhhDE, where h is any hydrophobic residue. The aspartic acid (D) helps coordinate the Mg^{2+} required for ATP hydrolysis and the glutamic acid residue (E) plays a role in activating water for the hydrolysis reaction [25,43,44]. Mutation of this latter residue prevents hydrolysis of the nucleotide, however does not prevent binding. In green-type activases, this same Walker A motif is present, however, the Walker B sequence is hhhhND, instead of the traditional sequence given prior. As with the glutamic acid residue, mutation of this aspartic acid residue in Rca leads to loss of the ability to hydrolyze the nucleotide, while still retaining the ability to bind it [45]. In addition, Sensors 1 and 2, seen in green and orange respectively, also play essential roles in coordinating and hydrolyzing bound nucleotides in activases [40,44]. Mutations in any of these two sequences seems to destabilize the binding of the nucleotide. A third feature, which helps with proper coordination of the nucleotide, and characteristic of AAA+ proteins, is the arginine finger. The finger is comprised of an Arg residue which helps generate the binding site by interacting with the adjacent subunit [44] and is crucial in nucleotide hydrolysis (**Fig. 1.7**). Despite having all

three activases belong to the same protein family, they are more different then they are alike; everything from their sequence identity, which is thought to be at most 25% over the

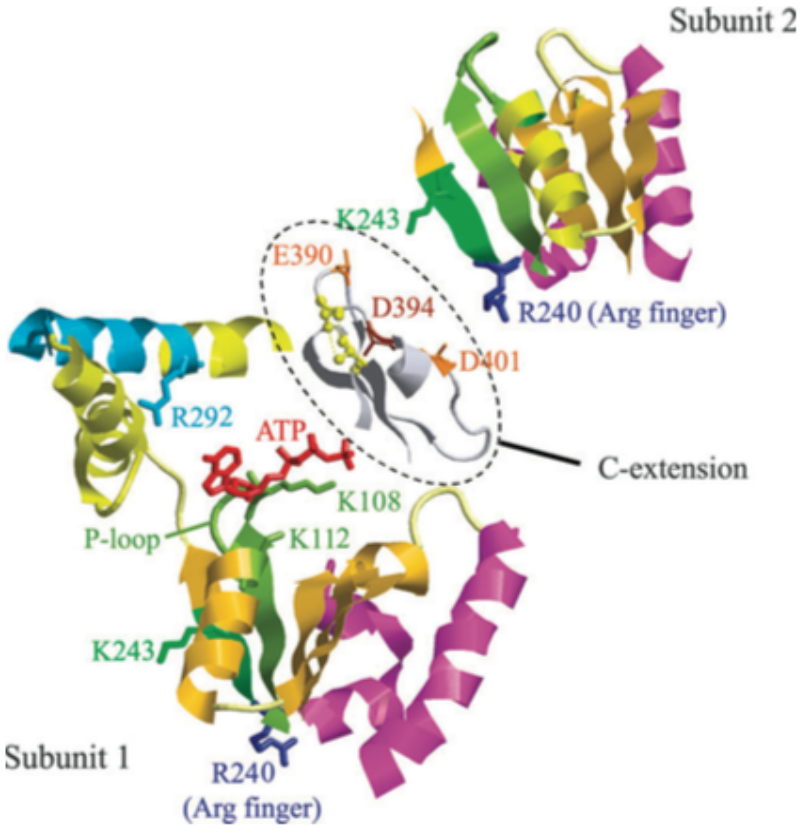


Figure 1.7: Regulatory role of C-terminal extension in long-form Rca. Proposed interaction and regulatory role of the C-terminal extension inside the binding site located at the interface of the two subunits. The structure shown is a model based on the known structure of Hslu (PDB code: 1G4l). This figure is adapted from (53).

AAA+ module, to their structural features, which have revealed substantial differences [40]. Therefore, despite being clustered together, each activases should be treated separately.

Red-type Activases:

Red-type activases derive their name from the fact that they reactivate form I Rubiscos from proteobacteria and members of the red plastid lineage. These activases are expressed on the same operon as the large and small subunits of Rubisco, causing scientists

to speculate that horizontal gene transfer was the mechanism involved in keeping all components together [25,40]. Structurally, red-type activases form obligate ring-like hexamers in the presence of RuBP and assemble from two different isoforms, one nuclear encoded isoform and one plastid encoded isoform, in an alternating pattern at a

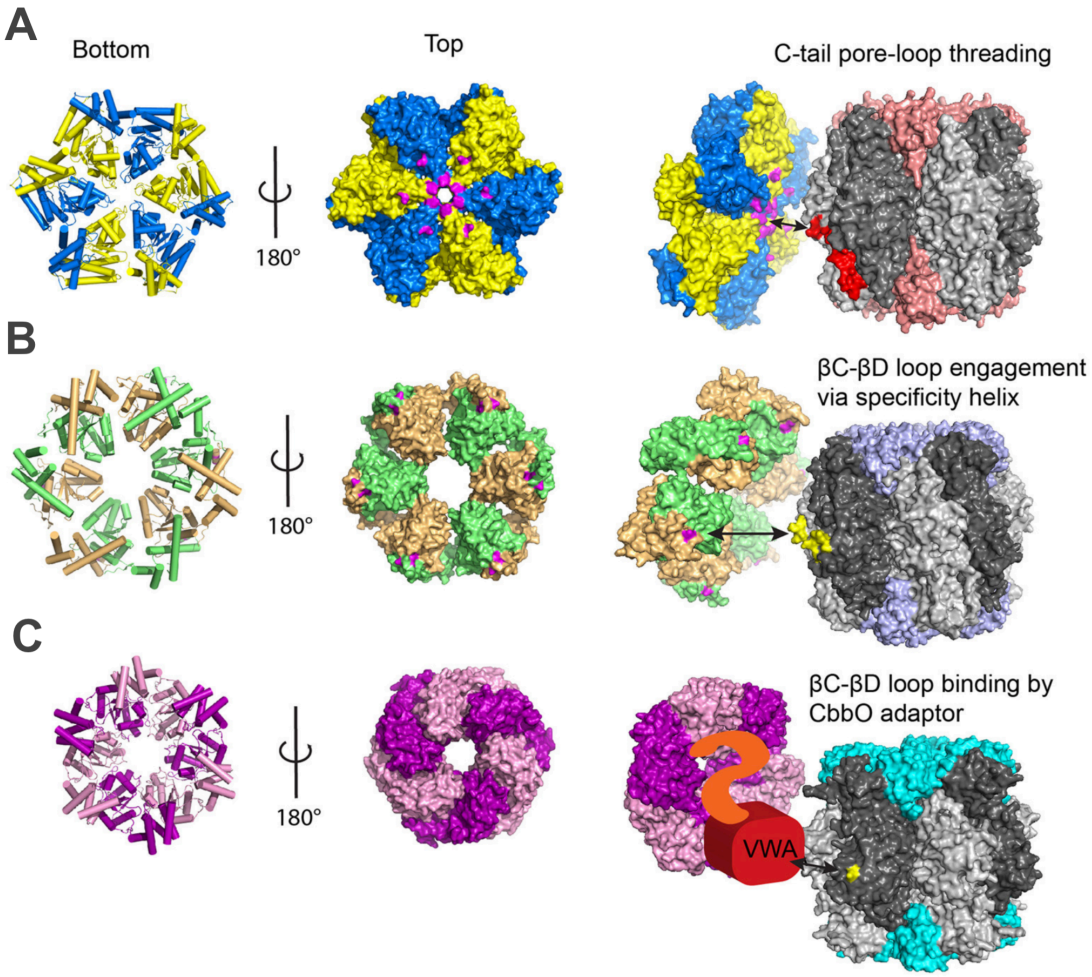


Figure 1.8: Structural and functional characterization of the Activase family. Structural views of either the (A) red-type activases, (B) green-type activases, or (C) CbbQO-type activases in either a bottom or top orientation. In addition, a general schematic representing their proposed reactivation mechanism is shown for all three activase types. This figure is adapted from (40).

stoichiometric ratio of 1:1 (**Fig. 1.8A**) [25,46,47]. The proposed mechanism for rubisco reactivation by red-type activases involves threading of Rubisco's C-terminal tail through the central pore of the closed ring hexamer, which results in a structural rearrangement of

Rubisco and causes release of RuBP [47,48]. This threading is powered by hydrolysis of ATP and is believed to be similar to mechanisms identified for other AAA+ proteins like ClpXP and HslU [49,50]. Once RuBP has been removed, the red-type activases release the tails and begin the process again.

CbbQO-type Activases:

Like some red-type activases, CbbQO-type activases are also found within proteobacteria, however, much less is known about them because they have been poorly characterized. They derive their unique name from their assembled state, which has multiple Q subunits assembling into a closed ring hexamer and an O subunit tethering itself to the ring in hinge-like fashion at a stoichiometric ratio of Q_6O_1 (**Fig. 1.8C**) [51,52]. The Q subunit is comprised primarily of the AAA+ domain while the O domain contains the von Willebrand Factor A (VWA) domain, a domain known to mitigate protein-protein interactions. Within the CbbQO family, there are different isoforms distinguished by their ability to reactivate different forms of Rubisco. The Q1O1 isoform is responsible for reactivating Form 1 Rubisco and the Q2O2 isoform is responsible for reactivating Form II Rubisco [40,51]. In both cases, reactivation of Rubisco is powered by ATP hydrolysis. The mechanism behind this reactivation is not quite understood in its entirety, but mutational analysis of the C-terminal tail of form I Rubisco has shown a significant decrease in reactivation by Q1O1, suggesting that some threading of the of Rubisco's C-terminal tail may be taking place. However, when mutations targeted at minimizing reactivation ability were made to the pore loop region of the Q1O1 assembly, reactivation activity was still observed, suggesting a secondary mechanism besides tail threading may be responsible [40]. Additionally, mutations to the MIDAS region of the VWA domain on the CbbO

subunit and to the conserved acidic acid residue between the βC-βD loop of the large subunit of Rubisco, resulted in almost complete loss of Rubisco reactivation, suggesting that the mechanism may be quite different than that seen in red-type activases (**Fig. 1.8C**) [40,51]. This same trend was also observed for the Q2O2 isoform.

Green-type Activase:

The green-type activases derive their name from the green lineage in which they are found. This lineage consists of higher plants and cyanobacteria [40]. Like other activases, green-type Rca have the canonical AAA+ domain, however, in addition, have extensions on both the N and C termini (**Fig. 1.6A**). Like the CbbQO-type activase, there are two different isoforms for the green-type activases, distinguished by the length of their C-terminal tails. The first, known as the α-isoform (also the long form), has a C-terminal extension containing redox-sensitive cysteine residues and several acidic ones [53]. These residues are believed to regulate the activation of Rubisco by regulating the binding of nucleotide to Rca [53,54]. For example, formation of a di-sulphide bond between two cysteines on the C-extension has been shown to lead to a conformational change that enables the entire extension to dock near, or in, the binding site [53,54] (**Fig. 1.7**). As one could image, this sort of structural rearrangement may negatively affect the relative activity of the activase [55]. The second isoform, known as the β-isoform (also the short form), lacks this redox-sensitive domain. In some species, like spinach, this shorter isoform can assemble together with the α-isoform and retain this method of regulation, however in others, like tobacco, only the β-isoform is present, preventing such regulation [53,56]. In addition, phosphorylation of certain residues may also serve to regulate Rca, despite not having the C-terminal extension [57].

There is currently a crystal structure of tobacco Rca, however there are a few serious limitations when analyzing it closely (**Fig. 1.8B**). The first is that the crystal structure is not complete. Only about 70% of the crystal structure has been solved, consisting of only the AAA+ module and lacking the N and C terminal ends, which are thought to play a substantial role in its function [42]. The second shortcoming is that the crystal structure is that of a spiraling assembly with six subunits per helical turn instead of a closed ring hexamer. However, due to the fact most AAA+ proteins form closed-ring hexamers [58,59], the Hayer-Hartl group used the density profile from a negative stain Electron Microscopy (EM) image of an assembly mutant containing a subunit interface substitution of R294V (tobacco numbering), and superimposed the crystal structure on it, generating the desired hexamer. This mutation disrupts the hydrogen bonding between the amide group on the Arg residue at position 294 (tobacco numbering) and the carboxyl group of the Asn residue at position 99 (tobacco numbering) of the adjacent subunit [42]. In trying to model it in this way, the authors had to group the α/β subdomain of one subunit with the α-helical subdomain of the adjacent subunit as rigid bodies and then tailor the hinge angles between the two subdomains of each subunit. This means that the hinge angle severely distorted, resulting in little to no density connecting the two subdomains of the same subunit [42]. The third limitation of this structure is that the crystal structure is of apo Rca, meaning that orientation and position of the nucleotide cannot be determined because it is not present. This last limitation is quite significant considering that a great deal of work was done to characterize the effect of nucleotide binding to the active site. In brief, work carried out by Hazra *et al*. showed that incubation with no nucleotide resulted in no cooperativity in ATPase activity (Hill coefficient = 1). However, addition of ADP seemed

to induce cooperativity (Hill coefficient = 1.9) suggesting that there may be major differences in conformations by having different nucleotides bound [60]. Additionally, varying ratios of ATP and ADP have also been shown to affect the reactivation activity and overall assembly behavior [56,61]. By not having a nucleotide bound structure, it is difficult to understand the implications of these different nucleotide occupancies on the overall fold and assembly of the protein.

A second structure of green-type Rca from *Arabidopsis thaliana* has been determined, but like the tobacco structure, it is not a closed-ring hexamer; it has the same spiraling structure and lacks the N and C termini (**Fig. 1.9**) [62]. However, instead of being Apo, the Arabidopsis structure does have sulfates bound in the active site, increasing the angle per helical turn and thus resulting in a more elongated structure (**Fig. 1.9**). Despite this improvement in hinge angle, great efforts are still underway to determine the structure of the full-length construct and have the nucleotide bound.

Functionally, green-type Rca plays a very important role in regulating the activity of form I Rubiscos. As mentioned above, Rubisco is prone to inhibition by a variety of different inhibitors, and therefore needs Rca to help remove the bound inhibitor. When looking at phenotypes, plants containing knockouts of the Rca gene showed very little growth, highlighting Rca's importance in plants [63]. The exact mechanism for reactivation of Rubisco is not entirely established, however, it is known that green-type Rca, like the other activases, hydrolyzes ATP in order to partially alter the structure of Rubisco, allowing the inhibitor to be removed [25,33]. It is also believed that green-type Rca does not thread the C-terminal tail of Rubisco through it's central pore, like red-type,

due to Rubisco's tail actually being about 10-12 residues shorter than Rubisco found in the red lineage [40]. Additionally, Rca can differentiate between Rubiscos from different

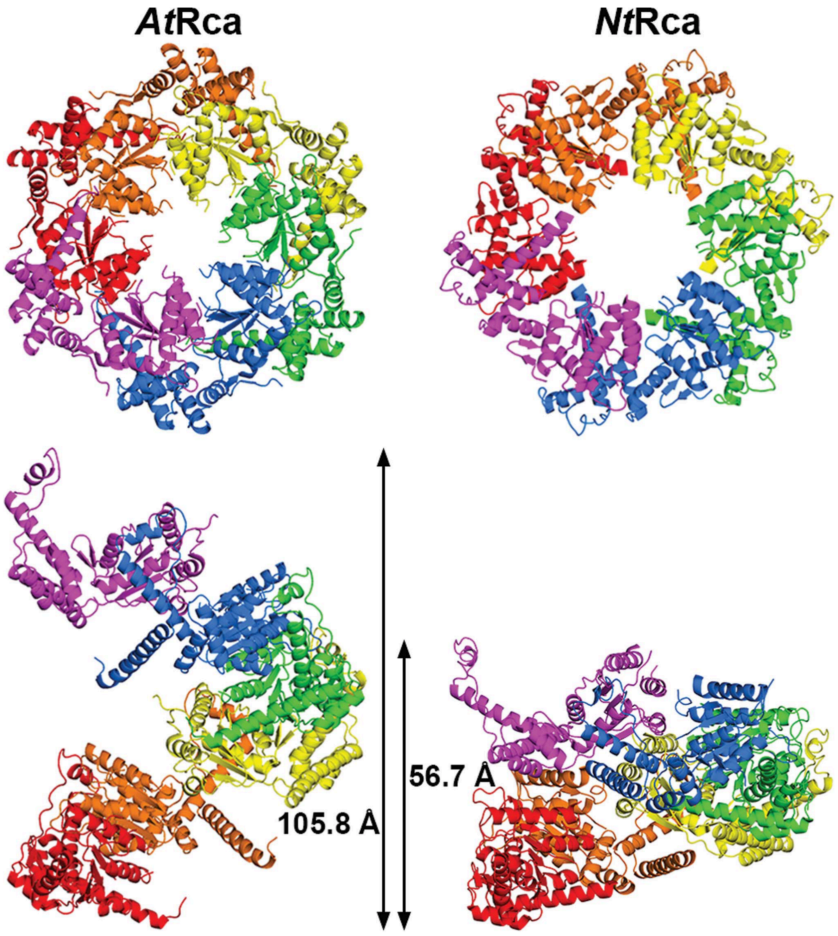


Figure 1.9: Crystal structure comparison between Rca from *Nicotiana tabacum* & *Arabidopsis thaliana*. Structural comparison between Rca from *Nicotiana tabacum* and *Arabidopsis thaliana* as seen from directly down the *z* axis (top panel) as well as perpendicular to the *z* axis (bottom panel). Measurements shown indicate the total distance spanning one helical turn containing 6 total subunits. This figure is adapted from (62).

species, using specific residues on it's H-9 helix [64,65], however, mutations to these residues can alter this specificity [66]. Given the fact that this interaction is so essential in regulation of Rubisco, and by extension essential in carbon fixation, much emphasis has

been placed on characterizing it functionally, even if the structural information is lacking.

The biggest hurdle in determining the mechanism is that unlike other activases and most AAA+ proteins, green-type Rca tends to be very polydisperse in its assembly, meaning that at any given concentration, there is a distribution of Rca oligomers rather than one single assembly [61,67,68]. Due to this polydisperse nature, traditional biochemical methods of assembly determination are rendered ineffectual because they are difficult to interpret. A great deal of time and energy was spent by many groups on developing and optimizing the methods for size-exclusion chromatography (SEC) [69] and Analytical Ultracentrifugation (AUC) [70,71], in hopes of determining the active oligomeric state. However, this work was ultimately unfruitful, generating more questions than answers, with the most significant limitation being the inability of these techniques to distinguish between active and inactive species. The Pearce group carried out a great deal of experiments comparing average molecular weights, measured as a function of concentration, with ATPase activity also collected as a function of concentration [70,71]. The result appeared to support a trimer as being the most active oligomer (given that the peak activity aligned well with an average MW corresponding to roughly a trimer). However, because the average MW is that of trimer, one could argue that a combination of dimers, tetramers, and hexamers could all be equally active as the weighted average of all three oligomers would result in the same average MW (**Fig. 1.10**). This finding from the Pearce group is in partial agreement with work carried out by the Hayer-Hartl group at the Max Planck Institute, which also found the most active oligomer in ATPase activity to be something between a dimer and tetramer [42]. However, in contrast, they found that the oligomer most active in Rubisco reactivation was something between a tetramer and

hexamer, making the interpretation less clear [42]. Therefore if we want to begin understanding activase function as it relates to assembly, then it is critical to know the

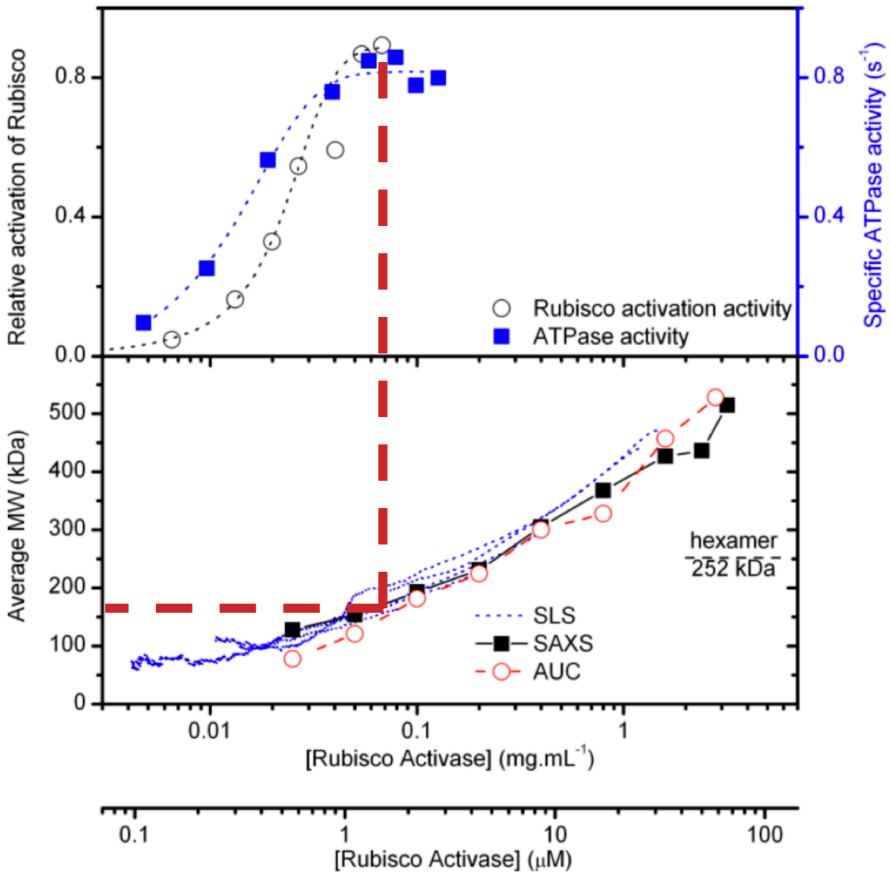


Figure 1.10: Identification of most active Rca oligomer as a function of concentration. Comparison between tobacco Rca activity, both in ATPase and Rubisco reactivation, compared to average MW extracted from SLS, SAXS, and AUC experiments as a function of concentration. Red dashed line serves as a visual guide for comparing maximum activity with corresponding MW. This figure is adapted from (70).

oligomeric distribution involved in reactivation of Rubisco.

To add to the complexity of higher plant Rcas, researchers have discovered that activity of Rca can vary from species to species and that there may be certain underlying assembly characteristics contributing to this variation in activity [70,72]. Work done in our lab, for example, has shown that the apparent melting temperature (T_m^{app}) of Rca, which

gives insight into Rca's thermal stability, varied from species to species. Cotton Rca, had a T_m value close to 42˚C where as spinach Rca, in identical conditions with ATP-γS, had a T_m^{app} value close to 58˚C [73]. This variation in T_m^{app} seemed to correlate with subunit exchange rates monitored by changes in FRET. When looking at the condition which gave the highest thermal stability in spinach Rca, Peterson-Forbrook *et al*. found that little to no subunit exchange was taking place, leading them to conclude that this high thermal stability is a result of the formation of tightly bound states that are essentially in a "locked" conformation [73]. Given that Rca has been shown to aggregate when subjected to non-ideal conditions, this variation in dynamics and thermal stability between different species could be due to drastic differences in assembly, which significantly affect activity, both ATPase and reactivation [74,75].

In addition to all reasons listed above, the last motivation for studying this self-assembly process is that without it, proper models trying to predict the interaction with Rubisco cannot be generated. Because Rca belongs to the AAA+ family, and because some data seems to suggest the formation of a hexamer [42,61,68)] models have been generated proposing two different methods of interaction. The first model, known as the "Top-on" model, shows the hexameric form of Rca interacting with the top of the Rubisco holoenzyme, particularly the small subunits, while the second model, known as the "Side-on" models, shows the hexameric Rca interacting primarily with the large subunits in the Rubisco holoenzyme [76]. The consensus is that the latter "Side-on" model is most consistent with the data, given that swapping of the small subunits with ones from other species caused little effect on activity [76].

Prior Assembly Work With Green-type Rca:

Thus far, a great deal of work has been done on characterizing the assembly of Rca as a function of concentration. Because traditional biochemical methods were difficult to

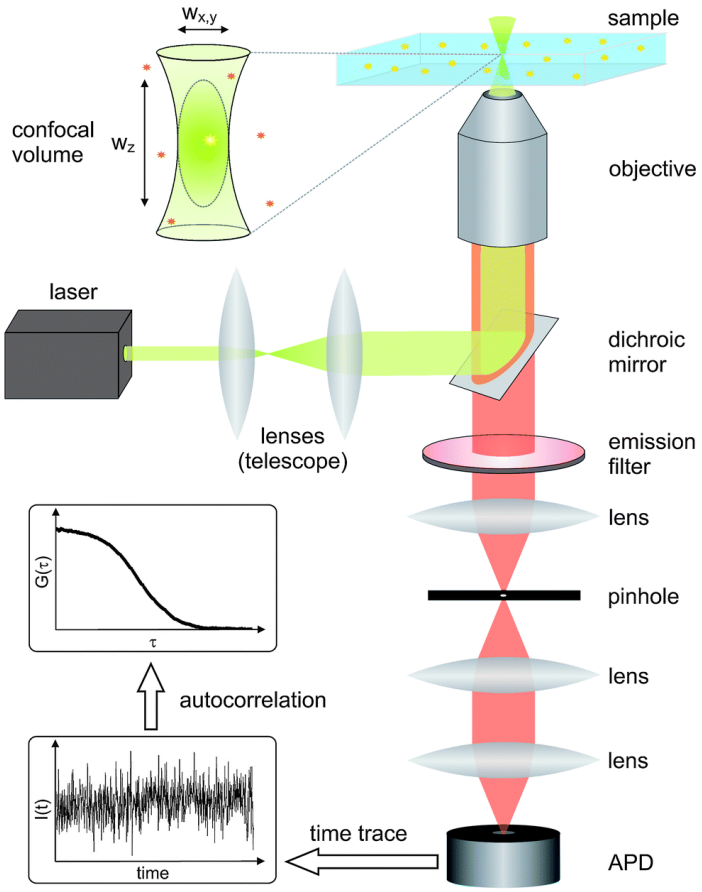


Figure 1.11: Depiction of homebuilt FCS setup. Cartoon visualization of a typical FCS set up. The light source strikes the sample via the objective, generating a fluorescent signal, which after hitting the APD, generates the fluorescence fluctuations, which are ultimately used to generate the autocorrelation decay. This figure is adapted from:
http://cam.facilities.northwestern.edu/588-2/fluorescence-correlation-spectroscopy/

interpret, the focus in the Wachter lab, in collaboration with the Levitus group, shifted to a fluorescence-based technique, specifically Fluorescence Correlation Spectroscopy (FCS), which provides greater sensitivity over bulk cuvette measurements, and has been used to address heterogeneity issues in other systems [77,78]. In brief, FCS works by generating a small confocal volume (approximately 7 fL) using optics, which is illuminated by a laser

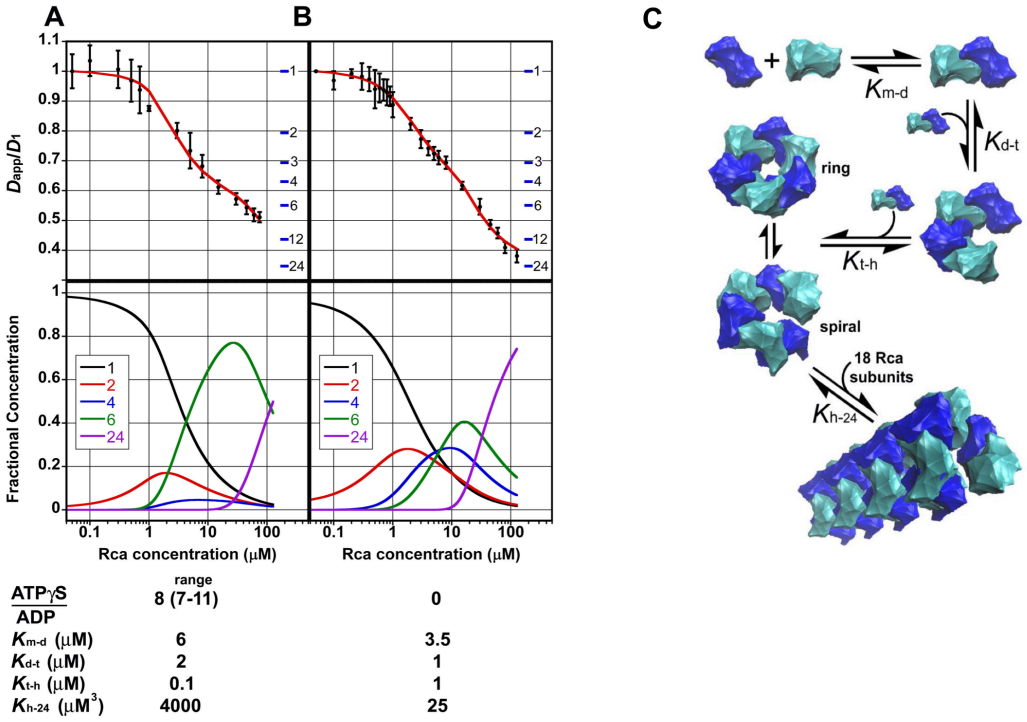
source. If the concentration of the fluorescently labeled sample is low enough (roughly 1 nM), then on average, 3 to 4 molecules of sample will be present inside the confocal volume at one time, thus making this a nearly single molecule technique [79]. Once the sample in the confocal volume gets illuminated, it fluoresces, generating a signal that is picked up by the detector, typically an Avalanche Photo Diode (APD) (**Fig. 1.11**). This signal generates fluctuations, which are used to generate an autocorrelation decay. This decay is then fit using the autocorrelation function, seen in equation 1 where G is the amplitude and r and z are the radial and axial semiaxes of the Gaussian confocal volume [67], to extract apparent diffusion coefficients (D_{app}) [79,80]. These D_{app} values can then be used to calculate the radius of the sample, using the Stokes-Einstein relation, seen in equation 2, where *k* is the Boltzmann constant, T is temperature, η is solvent viscosity, and r is the hydrodynamic radius.

$$G(\tau) = G_0 \left(1 + \frac{4D\tau}{{r_0}^2}\right)^{-1} \left(1 + \frac{4D\tau}{{z_0}^2}\right)^{-1/2} \qquad \text{Equation 1}$$

$$D = \frac{kT}{4\pi\eta r} \qquad \text{Equation 2}$$

Using this type of data processing, our lab was able to better characterize the distribution in oligomeric state of *Gossypium hirsutum* (cotton) Rca as a function of concentration, in the presence of both adenosine diphosphate (ADP) and ATP-γS, an non-hydrolyzable ATP analog (figure 6B). This work began by using an Rca construct

containing an engineered cysteine residue on the C-terminal tail that was solvent exposed and thus available for targeting by an Alexafluor 546 dye functionalized with a maleimide



ATPγS / ADP	range 8 (7-11)	0
K_{m-d} (μM)	6	3.5
K_{d-t} (μM)	2	1
K_{t-h} (μM)	0.1	1
K_{h-24} (μM^3)	4000	25

Figure 1.12: Assembly work done on Cotton short-form Rca. The assembly of Cotton Rca in the presence of **A**) ATP-γ-S, **B**) ADP, as a function of concentration (top panel) and fractional contribution of each oligomer at that respective concentration (bottom panel). The K_d values given below apply to the assembly model given in **C**. This figure is adapted from (61).

group. Once labeled, this sample was characterized by High Performance Liquid Chromatography (HPLC) and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) to ensure that there was only one label per monomer of Rca. With high quality samples ready, FCS measurements were collected as a function of concentration and normalized with respect to the monomeric assembly (**Fig. 1.12A and 1.12B Top panel**). This normalized data was then hand "fitted" using a mathematical model developed by our collaborators, allowing for input of different K_d

values based on a 1-2-4-6-24 assembly model. This model was selected over two other models, 1-6-24 and 1-3-6-24, because a dimeric species was clearly observed by SEC work [69]. In addition, this model was preferred because another AAA+ protein follows a similar assembly pathway which uses a dimer and tetramer intermediate before assembling into a hexamer [81]. Using this approach, we were able to extract fractional concentration profiles for each oligomer over the entire concentration range measured (**Fig. 1.12A and 1.12B Bottom panel**). From these plots, we observed that at low μM concentrations, Rca is primarily monomeric however quickly assembles into dimers and tetramers, which combine to generate hexamers. Despite this linear step in assembly, one finding that was clear is that assembly does not stop at the hexameric protomer, but rather continues to form higher ordered oligomers, termed twentyfourmers [67] (**Fig. 1.12C**). These large oligomers are believed to be spiraling species, consistent with the reported crystal structure. However, as is evident from FCS work, the hexamer is not always the predominant species. Additionally, work was done on a mutant lacking the ability to hydrolyze ATP, a mutant termed D173N [61]. More recent work carried out by the Moerner group, our collaborators at Stanford University, showed a high degree of similarity in this stepwise assembly process when examining the self-assembly pathway of Rca from *Nicotiana tabacum* (tobacco) on their ABEL Trap home-built instrument [82]. Like Cotton, assembly in the presence of ADP proceeded in a stepwise fashion towards larger oligomers, while assembly in the presence of ATP-γS appeared to favor more hexameric assemblies [55]. However, in contrast, Rca assembly in tobacco appeared to occur more rapidly, as larger oligomers were observed at lower concentrations [55].

Conclusions and Future Perspectives:

Despite having some insight into the complexity surrounding the activase family, the work on them is far from over. This is due in part to the fact that recent years have shown a drastic rise in interest within the biofuel and agricultural sectors focusing specifically on the overall increase in crop yield. By increasing these yields, we may potentially begin laying the groundwork for solving two major problems facing our every growing population, fuel and food. As mentioned above, recent trends suggest that supplies of food are not keeping in step with demands, which means the doors are wide open for innovators to come and begin solving the problem. Given that increases in yields depend primarily on a plant's ability to fix CO_2 and generate sugars for itself, work must begin by better understanding and characterizing the Rubisco/Rubisco activase interaction. Solving this issue will allow engineers to push this field forward and bring us back to production rate that can meet the ever-growing demand.

References:

1. Gust, D., *et al*. 2009. Solar Fuels via Artificial Photosynthesis. *Acc. Chem. Res*. **42**, 1890-1898

2. Zhu, X-G., *et al.* 2010. Improving Photosynthetic Efficiency for Greater Yield. *Annu. Rev. Plant Biol.* **61,** 235-261

3. Niinemets, U. 2007. Photosynthesis and resource distribution through plant canopies. *Plant Cell Environ.* **30,** 1052-1071

4. Hikosaka, K., and Terashima, I. 1995. A Model of the Acclimation of Photosynthesis in the Leaves of C-3 Plants to Sun and Shade with Respect to Nitrogen use. *Plant Cell Environ.* **18,** 605-618

5. Peng, S. B., *et al.* 2009. Current Status and Challenges of Rice Production in China. *Plant Prod. Sci.* **12,** 3-8

6. Wiebe K, *et al*. 2008. The State of Food and Agriculture 2008. Biofuels: Prospects, Risks and Opportunities. Food and Agriculture Organization of the United Nations, Rome, Italy

7. Parry, M. A., *et al*. 2013. Rubisco Activity and Regulation as Target for Crop Improvements. *J. Exp. Biol*. **64,** 717-730

8. Vinyard, D. J., *et al*. 2013. Photosystem II: The Reaction Center of Oxygenic Photosynthesis. *Annu. Rev. Biochem*. **82,** 577-606

9. Dekker, J.P., and Boekema, E.J. 2005. Supramolecular Organization of Thylakoid Membrane Proteins in Green Plants. *Biochim. Biophys. Acta.* **1706,** 12–39

10. Schubert, W. D., *et al.* 1998. A Common Ancestor for Oxygenic and Anoxygenic Photosynthetic Systems: A Comparison Based on the Structural Model Of Photosystem I. *J. Mol. Biol.* **280,** 297–314

11. Allen, J. F. 2002. Photosynthesis of ATP – Electrons, Proton Pumps, Rotors, and Poise. *Cell.* **110,** 273-276

12. Umena, Y., *et al.* 2011. Crystal Structure of Oxygen-Evolving Photosystem II at a Resolution of 1.9 Å. *Nature.* **473,** 55–60

13. Coe, J., et al. 2015. Crystallization of Photosystem II for Time-Resolved Structural Studies Using an X-ray Free Electron Laser. *Methods Enzymol.* **557,** 459-482

14. Kirchhoff, H., *et al.* 2017. Sublocalization of Cytochrome b6f Complexes in Photosynthetic Membranes. *Trends Plant Sci.* **22,** 574-582

15. Baniulis, D., *et al*. 2008. Structure–Function of the Cytochrome b6f Complex. *Photochem. Photobiol*. **84,** 1349–1358

16. Fromme, P., *et al.* 2001. Structure of Photosystem I. Biochim. *Biophys. Acta*. **1507,** 5–31

17. Gisriel, C., *et al.* 2017. Structure of a Symmetric Photosynthetic Reaction Center–Photosystem. *Science.* **357**, 1021–1025

18. Junge, W., and Nelson, N. 2015. ATP Synthase. *Annu. Rev. Biochem*. **84,** 631-657

19. Sambongi, Y., *et al.* 1999. Mechanical Rotation of the C-Subunit Oligomer in ATP Synthase (FOF1): Direct Observation. *Science* **286,**1722–1724

20. Woodrow, I. E., and Berry, J. A. 1988. Enzymatic Regulation of Photosynthetic CO2 Fixation in C3 Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **39,** 533-594

21. Andersson, I. and Backlund, A. 2008. Structure and Function of Rubisco. *Plant Physiol. Biochem.* **46,** 275-291

22. Shimizu, R. *et al*. 2015. New Insight Into the Role of the Calvin Cycle: Reutilization of CO2 Emitted Through Sugar Degradation. *Scientific Reports*. **5,** 11617. DOI: 10.1038/srep11617

23. Whitney, S.M., *et al.* 2011 Advancing our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco. *Plant Physiol.* **155,** 27-35

24. Tabita, F. R. *et al*. 2008. Distinct Form I, II, III, And IV Rubisco Proteins From the Three Kingdoms of Life Provide Clues About Rubisco Evolution and Structure/Function Relationships. *J. Exp. Bot.* **59,** 1515-1524

25. Mueller-Cajar, O., *et al.* 2013. Maintaining Photosynthetic CO2 Fixation via Protein Remodeling: The Rubisco Activases. *Photosynth. Res.* **DOI:** 10.1007/s11120-013-9819-0

26. Cleland, W.W., *et al.* 1998. Mechanism of Rubisco: The Carbamate as a General Base. *Chem. Rev.* **98,** 549-561

27. Bauwe, H., et al. 2010. Photorespiration: Players, Partners and Origin. *Trends Plant Sci.* **15,** 330-336

28. Schwarte, S., and Bauwe, H. 2007. Identification of the Photorespiratory 2-Phosphoglycolate Phosphatase, PGLP1, in Arabidopsis. *Plant Physiol.* **144,** 1580-1586

29. Bracher, A., *et al.* 2015. Degradation of Potent Rubisco Inhibitor by Selective Sugar Phosphate. *Nat. Plants* **1,** 14002. DOI: 10.1038/NPLANTS.2014.2

30. Parry, M. A. J., *et al.* 2008. Rubisco Regulation: A Role for Inhibitors. *J. Exp. Bot*. **59,** 1569-1580

31. Andralojc, P. J. *et al.* 2012. 2-carboxy-D-arabinitol 1-Phosphate (CA1P) Phosphatase: Evidence for a Wider Role in Plant Rubisco Regulation. *Biochemistry* **442**, 733-742

32. Andersson, I. 2008. Catalysis and Regulation in Rubisco. *J. Exp. Bot*. **59,** 1555-1568

33. Wachter, R. M. and Henderson, J. N. 2015. Rubisco Rescue. *Nature Plants* **1,** 1-2

34. Mackinder, L. C. M., *et al*. 2016. A Repeat Protein Links Rubisco to form the Eukaryotic Carbon-Concentrating Organelle. *Proc. Natl. Acad. Sci*. **113**, 5958-5963

35. Gunn, L. H., *et al*. 2017. A unique structural domain in Methanococcoides burtonii ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) acts as a small subunit mimic. *J. Biol. Chem*. **292,** 6838-6850

36. Tcherkez, G. G. B., *et al*. 2006. Despite Slow Catalysis and Confused Substrate Specificity, all Ribulose Bisphosphate Carboxylases May be Nearly Perfectly Optimized. *Proc. Natl. Acad. Sci*. **103,** 7246-7251

37. Parry, M.A.J., *et al*. 2003. Manipulation of Rubisco: The Amount, Activity, Function and Regulation. *J. Exp. Bot.* **54,** 1321-1333

38. Aigner, H. *et al*. 2017. Plant RuBisCo Assembly in E. coli with Five Chloroplast Chaperones Including BSD2. *Science.* **358,** 1272-1278

39. Durão, P. *et al.* 2015. Opposing Effects of Folding and Assembly Chaperones on Evolvability of Rubisco. *Nat. Chem. Biol.* **11,** 148-155

40. Mueller-Cajar, O. 2017. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Front Mol. Biosci.* **4**, 31. DOI: 10.3389/fmolb.2017.00031

41. Portis, Jr. A. R. and Salvucci, M. E. 2002. The Discovery of Rubisco Activase – Yet Another Story of Serendipity. *Photosynth. Res.* **73,** 257-264

42. Stotz, M., *et al*. 2011. Structure of Green-type Rubisco activase from Tobacco. *Nat. Struct. Mol. Biol.* **18,** 1366-1370.

43. Wendler, P. *et al*. 2012. Structure and Function of the AAA+ Nucleotide Binding Pocket. *Biochim. Biophys. Acta*. **1823,** 2-14

44. Hanson, P. I. and Whiteheart, S. W. AAA+ Proteins: Have Engine, Will Work. *Nat. Rev. Mol. Cell Biol.* **6,** 519-529

45. Van de Loo, F. J., and Salvucci, M. E. 1998. Involvement of Two Aspartate Residues of Rubisco Activase in Coordination of the ATP γ-Phosphate and Subunit Cooperativity. *Biochemistry* **37,** 4621-4625.

46. Mueller-Cajar, O., *et al*. 2011. Structure and Function of the AAA+ Protein Cbbx, a Red-Type Rubisco Activase. *Nature* **479,** 194-199

47. Loganathan, N., *et al.* 2016. Characterization of the Heterooligomeric Red-Type Rubisco Activase from Red Algae. *Proc. Natl. Acad. Sci.* **113,** 14019-14024

48. Bhat, J. Y., *et al.* 2017. Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Mol. Cell* **67,** 1-13

49. Rodriguez-Aliaga, P., *et al.* 2016. Substrate-Translocating Loops Regulate Mechanochemical Coupling and Power Production in AAA+ Protease Clpxp. *Nat. Struct. Mol. Biol.* **23,** 974-981

50. Baytshtok, V., *et al.* 2016. A Structurally Dynamic Region of the HslU Intermediate Domain Controls Protein Degradation and ATP Hydrolysis. *Structure* **24,** 1-12

51. Tsai, Y. C., *et al.* 2015. Identification and Characterization of Multiple Rubisco Activases in Chemoautotrophic Bacteria. *Nat. Commun.* **6,** 8883. DOI:10.1038/ncomms9883

52. Sutter, M. *et al.* 2015. Structural Characterization of a Newly Identified Component of α-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Scientific Reports.* **5,** 16243. DOI: 10.1038/srep16243

53. Portis, Jr. A. R., *et al.* 2008. Regulation of Rubisco activase and its Interaction with Rubisco. *J. Exp. Bot.* **59,** 1597-1604.

54. Zhang, N., *et al*. 2001. Characterization of the Regulatory Function of the 46-kD Isoform of Rubisco activase from Arabidopsis. *Photosynth. Res.* **68,** 29-37.

55. Wang, D., and Portis Jr., A. R. 2006. Increased Sensitivity of Oxidized Large Isoform of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase to ADP Inhibition Is Due to an Interaction between its Carboxyl Extension and Nucleotide-binding Pocket. *J. Biol. Chem.* **281,** 25241-25249

56. Carmo-Silva, A. E. and Salvucci, M. E. 2013. The Regulatory Properties of Rubisco Activase Differ Among Species and Affect Photosynthetic Induction During Light Transitions. *Plant Physiol.* **161,** 1645-1655

57. Boex-Fontvieille, E. *et al.* 2014. Phosphorylation Pattern of Rubisco Activase in Arabidopsis Leaves. *Plant Biol.* **16,** 550-557

58. Schuller, J.M., *et al.* 2016. Nucleotide-Dependent Conformational Changes of the AAA+ ATPase p97 Revisited. *FEBS Letters* **DOI:**10.1002/1873-3468.12091

59. Su, S-C., *et al.* 2016. Structural Basis for the Magnesium-Dependent Activation and Hexamerization of the Lon AAA+ Protease. *Structure.* **24,** 676-686

60. Hazra, S., *et al.* 2015. Regulation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase: Product Inhibition, Cooperativity, and Magnesium Activation. *J. Biol. Chem.* **290,** 24222-24236

61. Kuriata, A. M. *et al.* 2014. ATP and Magnesium Promote Cotton Short-Form Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Hexamer Formation at Low Micromolar Concentrations. *Biochemistry* **53,** 7232-7246.

62. Hasse, D., *et al.* 2015. Structure of Arabidopsis thaliana Rubisco activase. *Acta Cryst.* **D71,** 800-808. DOI: 10.1107/S1399004715001182

63. Kurek, I. *et al.* 2007. Enhanced Thermostability of Arabidopsis Rubisco Activase Improves Photosynthesis and Growth Rates under Moderate Heat Stress. *Plant Cell* **19,** 3230-3241

64. Li, C., *et al.* 2005. Two Residues of Rubisco Activase Involved in Recognition of the Rubisco Substrate. *J. Biol. Chem.* **280,** 24864-24869

65. Henderson, J. N., *et al.* 2011. Atomic Resolution X-ray Structure of the Substrate Recognition Domain of Higher Plant Rubisco Activase. *J. Biol. Chem*. **286,** 35683-35688.

66. Ott, C. M., *et al*. 2000. Activase Region on Chloroplast Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase - Nonconservative Substitution in the Large Subunit Alters Species Specificity of Protein Interaction. *J. Biol. Chem*. **275,** 26241-26244.

67. Chakraborty, M. *et al.* 2012. Protein Oligomerization Monitored by Fluorescence Fluctuation Spectroscopy: Self-Assembly of Rubisco activase. *Biophys. J.* **103,** 949-958.

68. Wang, Q., *et al.* 2018. Single-molecule diffusometry reveals the nucleotide-dependent oligomerization pathways of Nicotiana tabacum Rubisco activase. *J. Chem. Phys*. **148,** 123319. DOI: 10.1063/1.5005930

69. Henderson, J. N., *et al.* 2013. Biophysical Characterization of Higher Plant Rubisco Activase. *Biochim. Biophys. Acta*. **1834,** 87-97

70. Keown, J. R., *et al.* 2013. Small Oligomers of Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase are Required for Biological Activity. *J. Biol. Chem.* **288,** 20607-20615.

71. Keown, J. R. and Pearce, G. 2014. Characterization of Spinach Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Activase Isoforms Reveals Hexameric Assemblies with Increased Thermal Stability. *Biophys. J.* **464,** 413-423.

72. Carmo-Silva, A. E., and Salvucci, M.E. 2011. The Activity of Rubisco's Molecular Chaperone, Rubisco Activase, in Leaf Extracts. *Photosynth. Res.* **108,** 143-155.

73. Peterson-Forbrook, D. S., *et al.* 2017. Nucleotide Dependence of Subunit Rearrangements in Short-Form Rubisco Activase from Spinach. *Biochemistry* **56,** 4906-4921

74. Barta, C. *et al.* 2010. Structural Changes Associated with the Acute Thermal Instability of Rubisco Activase. Arch. *Biochem. Biophys.* **499,** 17-25

75. Scafaro, A. P., *et al.* 2016. Heat Tolerance in a Wild Oryza Species is Attributed to Maintenance of Rubisco Activation by a Thermally Stable Rubisco Activase Ortholog. *New Phytol.* **211,** 899-911

76. Wachter, R. M., *et al*. 2013. Activation of Interspeceies-hybrid Rubisco Enzymes to Assess Different Models for the Rubisco-Rubisco activase Interaction. *Photosynth. Res.* **117,** 557-566.

77. Sengupta, P., *et al.* 2003. Measuring Size Distribution in Highly Heterogeneous Systems with Fluorescence Correlation Spectroscopy. *Biophys. J.* **84,** 1977-1984

78. Módos, K., *et al.* 2004. Maximum-Entropy Decomposition of Fluorescence Correlation Spectroscopy Data: Application to Liposome-Human Serum Albumin Association. *Euro. Biophys. J.* **33,** 59-67

79. Lakowicz, J. R. 2006. Principles of Fluorescence Spectroscopy. Springer 3rd Edition, 798-840.

80. Jameson, D. M. 2014. Introduction to Fluorescence. CRC Press 1st Edition, 163-190.

81. Kress, W., *et al.* 2007. Assembly Pathway of an AAA+ Protein: Tracking ClpA and ClpAP Complex Formation in Real Time. *Biochemistry* **46,** 6183-6193

82. Wang, W. and Moerner, W. E. 2014. Single-molecule Motions Enable Direct Visualization of Biomolecular Interactions in Solution. *Nat. Methods* **11,** 555-558

CHAPTER 2

REGULATION OF ATPASE ACTIVITY BY DYNAMIC SELF-ASSEMBLY IN TOBACCO RCA

Abstract:

Rubisco activase (Rca) from higher plants is a stromal ATPase essential for reactivating Rubiscos rendered catalytically inactive by endogenous inhibitors. Rca's functional state is thought to consist of ring-like hexameric assemblies, similar to other members of the AAA+ protein superfamily. However, unlike other members, it does not form obligate hexamers and is quite polydisperse in solution, making elucidation of its self-association pathway challenging. This polydispersity also makes interpretation of traditional biochemical approaches difficult, prompting us to use a fluorescence based technique (Fluorescence Correlation Spectroscopy) to investigate the relationship between quaternary structure and function. Like cotton β Rca, tobacco β Rca appears to assemble in a step-wise and nucleotide dependent manner, following a similar aggregate model. High concentrations of ADP, as well as the absence of nucleotides, seems to favor continuous assembly towards larger oligomers, while assembly in the presence of ATP-γS (an ATP analog) appears to halt continuous assembly in favor of hexameric species. A common element in all three, however, is a significant minimization of the tetrameric species. In contrast, assembly in the "Active ATP Turnover" condition (a mixture of ATP and ADP) appears to favor an almost equal distribution of tetramer and hexamer. Additionally, varying the nucleotide ratio up or down, positively or negatively effects the tetramer distribution, suggesting that a mixture of nucleotides is necessary for tetramer accumulation. The importance of the tetramer accumulation is further made evident by a

comparison with activity assays monitoring ATP turnover of Rca as a function of concentration, which show maximum activity in the low μM range (between 0.5 and 2.5 μM) corresponding to a tetramer/hexamer ratio of roughly 1. Additionally, the observed decrease in ATPase activity with increasing protein concentration, does not follow the decrease of any particular oligomer, but rather decreases with increasing aggregation, suggesting that assembly dynamics may regulate ATPase activity, rather than the formation/disappearance of one specific oligomer. Like wild-type, assembly and activity characterization of the subunit-subunit interface mutant R294V, showed similar correlation between maximum ATPase activity and nearly equal tetramer/hexamer accumulation, further supporting the idea of regulation by assembly. Given this relationship, one could argue that assembly dynamics are essential for Rca function, both in ATPase and in regulation of Rubisco carboxylation activity, thus providing a rational for Rca's high degree of polydispersity.

Introduction:

The photosynthetic enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) carries out the vast majority of CO_2 fixation reactions on earth. In part, the activity of Rubisco is regulated by the binding of competitive inhibitors, such as phosphorylated metabolites that are at times generated during turnover [1] or various endogenous inhibitors [2]. Under some conditions, the substrate ribulose-1,5-bisphosphate (RuBP) itself may serve as a tight-binding inhibitor. Critical in Rubisco regulation is the AAA+ protein Rubisco activase (Rca), an ATPase that speeds up inhibitor release by mechanically adjusting the conformation of Rubisco [3]. Under high irradiance, Rca provides continuous support of Rubisco activity, whereas under rapidly changing light conditions, Rca ensures the proper alignment of the light and dark reactions of photosynthesis [4]. All known proteins with Rca activity are members of the AAA+ superfamily, and appear to have arisen by convergent evolution in many photosynthetic organisms [5-8]. All form multi-subunit assemblies from two-domain protomers that bear a nucleotide binding pocket near a flexible hinge region.

In addition to the AAA+ module, higher plant Rca proteins are extended N-terminally by about 68 residues, and C-terminally by about 19 residues for short-form (β) Rca, and about 49 residues for long-form (α) Rca. Tobacco expresses only β-Rca, a 383 residue protein, missing the longer isoform that appears to be redox-regulated by electron transport through the thylakoid membrane [9,10]. Under some conditions, tobacco Rca has been shown to form hexameric ring-like assemblies [11], although in general, preparations of plant Rca are known to be highly polydisperse [12,13]. In the X-ray structure of the apo-AAA+ domain of tobacco and Arabidopsis Rca, a helical assembly was observed that

contained six subunits per turn [11,14]. Structural information for the N- and C-terminal extensions remains sparse, as these segments were either disordered or absent in the crystals.

In tobacco Rca, the R294V substitution has been used extensively to investigate quaternary structural features. Arg294 is found in the AAA+ sensor II domain, a region that contains several conserved arginine residues that may interact with the γ-phosphoryl group of bound ATP [15]. Based on the X-ray structure of the tobacco AAA+ domain [11], replacement of Arg294 with a valine residue is expected to disrupt a hydrogen bond between the guanidinium group of Arg294 and the amide nitrogen of Asn99 of the adjacent subunit. Presumably, the effect would be a loosening of subunit-subunit interactions, as Arg294 would not be able to reach the nucleotide binding pocket [16], although its replacement may affect hinge-bending motions of the AAA+ domain, and therefore packing of one subunit against the other.

Negative-stain electron microscopic (EM) images collected on 500 nM tobacco Rca-R294V provided evidence of closed hexameric rings in the presence of ATP-γS. Surprisingly, the homologous substitution introduced into *Gh* β-Rca was reported to render this protein incapable of nucleotide binding [13]. The EM electron density was used as a basis for the generation of a closed-ring model of the AAA+ domain from the tobacco crystal structure [11].

Experiments in mechanistic enzymology carried out on tobacco Rca indicated that positive cooperativity (Hill coefficient of 1.9) requires both ATP- and ADP-bound subunits [17]. Occupation of a small fraction of sites by ADP was shown to be coupled to high turnover rates [17]. In the presence of excess magnesium, the K_m^{ATP} was estimated to be

104 μM, about three times as high as the K_i^{ADP} of 37 μM. The existence of three different classes of sites in multimeric assemblies is experimentally supported by the Hill coefficient for magnesium activation, n_H^{Mg} = 2.5 [17]. Activation of hydrolysis of the substrate Mg·ATP is thought to involve the binding of a second Mg^{2+} at the subunit-subunit interface in the vicinity of the active site [13]. Recently, we have demonstrated that a Rubisco:Rca_6 ratio as high as 24:1 does not alter Rca hydrolytic activity, not even in the presence of crowding agents [17]. The complete lack of catalytic stimulation by inhibited Rubisco has remained somewhat perplexing [11,17].

The interactions of Rca with ATP analogs have been characterized to various degrees. For cotton-β-Rca, low turnover of ATP-γS has been reported [13] in conjunction with a small elevation in apparent melting temperature T_m^{app} from ~42°C (Mg·ATP) to 43°C (Mg·ATP-γS) [18]. To the contrary, with spinach Rca, a tremendous jump in thermal stabilization has been observed upon binding of ATP-γS [18,19]. The T_m^{app} was shown to rise from 35°C with Mg·ATP to 60°C with Mg·ATP-γS. In the presence of the transition state analog ADP·Al(Fx), the T_m^{app} was shown to rise to 49°C [18]. In spinach β-Rca, ATP-γS exhibits a $K_i^{ATP\text{-}\gamma S}$ about 10-fold lower than the Michaelis constant K_m^{ATP}, and is hydrolyzed about 40-fold slower than ATP [18]. Subunit exchange is arrested in the presence of either ATP-γS or ADP·Al(Fx), in line with a tight-binding pre-hydrolysis state similar to the locked state that has been characterized in the bacterial AAA+ protease ClpX [20]. In the presence of ATP, subunit exchange of spinach β-Rca proceeds only slowly until ADP has accumulated to a ratio of ATP/ADP = 4:1, suggesting that the exchange rate

is a function of the distribution of tight- and loose-binding conformations in oligomeric assemblies [18].

Over the years, the stabilization of a hexameric state by ATP-γS has been reported for tobacco Rca [11,21] and cotton β-Rca [13]. In 2012, we established fluorescence correlation spectroscopy (FCS) as a means to estimate the oligomeric composition of Rca preparations as a function of subunit concentration [12]. Our work on cotton β-Rca demonstrated that hexamerization involves well-populated intermediate states [12], leading to an assembly pathway involving dimers and tetramers, as later verified by single molecule diffusometry [21]. In this work, the continued assembly from hexameric to higher order states was mathematically modeled by the association of four hexamers to form a 24-mer. Although populated, intermediates in this range were not explicitly represented. In a follow-up study also on cotton β-Rca, we demonstrated that higher ATP/ADP ratios promoted hexamer accumulation [13], suggesting that a closed-ring toroidal form may be favored under these conditions. On the other hand, ADP as the sole nucleotide was proposed to favor open "lock washer"-type hexameric structures that could serve as templates for continued assembly to supramolecular spirals. In the same work, we demonstrated that high magnesium plays a central role in hexamer formation [13].

Sedimentation velocity experiments on wild-type tobacco Rca indicated that within a concentration range of 1 to 20 μM, the protein assembles continuously to form larger and larger species [19]. The molecular mass appeared to range from trimers to dodecamers, and bound nucleotide (ADP, ATP-γS, apo-protein) was reported to have no effect on average particle size. Surprisingly, maximal ATPase activity was observed at 1 μM Rca, where assemblies were estimated to consist of about 3 subunits [19]. The minimal

stoichiometry for ATP hydrolysis has also been estimated by use of unidirectional subunit interface mutants of tobacco Rca [11]. Titration experiments suggested that the smallest oligomer capable of ATP hydrolysis consists of about three subunits.

The interface mutant R294V was also investigated by analytical ultracentrifugation (AUC) [19]. In the presence of ATP-γS, the data were reported to be consistent with a hexameric species between 1 and 20 μM. In contrast, the apo-form of tobacco Rca-R294V appeared to assemble more slowly, whereas ADP induced the formation of supramolecular aggregates [19].

Recently, single-molecule diffusometry has provided substantial insight into the early steps of the tobacco Rca assembly process. Using an aqueous phase Anti-Brownian Electrokinetic (ABEL) trap, heterogeneous, transient oligomer distributions could be observed directly, and reliable K_d values for the monomer-dimer and dimer-tetramer equilibria could be extracted [21]. Although the method did not allow for a distinction between tetramers and hexamers, the results clearly demonstrated that the assembly of tobacco Rca does depend on the type of nucleotide, and that ATP-γS favors the formation of tetramer/hexamer populations at the expense of dimers and larger aggregates. In addition, the rate of subunit exchange involving dimers and tetramer/hexamer populations was shown to be about 0.2 s^{-1} with Mg·ADP and 0.1 s^{-1} with Mg·ATP-γS [21].

In the present work, we use FCS to demonstrate that tobacco Rca needs a mix of nucleotides for tetramer accumulation. Further, we find that the ATPase activity reaches a maximum at a concentration with substantial amounts of dimer, tetramer and hexamer. Our data are most consistent with a model that couples hexameric ATPase activity with its dissociation to a dimer and tetramer.

Materials and Methods:

Plasmid generation and mutagenesis. The gene encoding for Nt-β-Rca, residues 1-383, was removed from the pET23a(+) vector and inserted into the pHUE expression system, using the SacII and NotI restriction sites. This expression system was a gift from S. Whitney (Australian National University, Canberra, Australia) and allows for fusion of Nt-β-Rca with a 6His-tagged ubiquitin tag, which can be cleaved by a 6His-tagged deubiquitylating enzyme (DUB).

In order to specifically label the protein with a fluorescent dye, a cysteine residue was introduced at position 379 by replacing an existing serine (S379C) using the QuikChange site-directed mutagenesis kit (Agilent) and provided instructions. Additionally, a second mutation was introduced at position 294 by replacing arginine with valine (R294V), to generate the assembly mutant identified in the literature. The entire sequence for both mutant constructs were verified by DNA sequencing. Full-length Nt-β-Rca carrying the single cysteine substitution was denoted "Wild-type" while that carrying the double substitution was designated Assembly mutant.

Transformation and Growth. The pHUE vector containing either the "Wild-type" or Assembly mutant gene was transformed into *Escherichia coli* strain BL21*(DE3) (Invitrogen) using the heat-shock method. Single colonies were picked and grown overnight (14 – 16 hours) in LB media containing 100 μg/mL carbenicillin. This overnight culture was then used to inoculate a (6) 1L culture of LB media also containing 100 μg/mL carbenicillin. Inoculated cultures were grown at 37 ˚C and 250 rpm, until the OD_{600} reached 0.6, then induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

After induction, the cultures were grown for 8 – 10 hours at 25 ˚C and 200 rpm. Cells were then harvested via centrifugation (6K rpm for 20 min at 4 ˚C) and frozen at -80 ˚C.

Protein Purification. The cell pellet was suspended in 100 mL of 25 mM Tris-HCl (pH 8.0), 10 mM Imidazole (pH 8.0), 30% glycerol, 150 mM NaCl, 0.1 mM EDTA (pH 8.0), 1 mM PMSF, and 30 mg of hen egg white lysozyme and stirred for 15 minutes at 4 ˚C. Cells were then lysed by 5 rounds of sonication (60 s on, 60 s off), and centrifuged at 15K rpm for 30 minutes at 4˚C. The supernant was then passed through 0.22 μm syringe filter and loaded onto a HisTrap HP nickel-nitrilotriacetic acid (Ni-NTA) column (GE) on the ÄKTA pure FPLC system (GE). The HisTrap HP column was then washed with 10 mM Imidazole (Solvent A1: 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM Imidazole pH 8.0, 0.5 mM ADP and 1 mM DTT) to eliminate unbound species. A step gradient to 100 mM Imidazole (Solvent B: 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 500 mM Imidazole pH 8.0, 0.5 mM ADP and 1 mM DTT) was then used to wash off weakly bound species. Desired Nt-β-Rca was eluted with a linear gradient from 100 mM to 500 mM Imidazole (Rca eluted at 200 mM Imidazole). After pooling all fractions containing Nt-β-Rca, 0.4 mg of deubiquitylating enzyme (DUB) (final concentration of 11.4 μg/mL or 0.27 μM) was added to the sample and was dialyzed overnight at 4 ˚C against 1 L of Dialysis buffer (25 mM Tris-HCl pH 8.0, 40 mM NaCl, 0.1 mM EDTA pH 8.0, 0.5 mM ADP and 1 mM DTT). After dialysis, the sample was reapplied to the HisTrap HP column and washed with 10 mM Imidazole (Solvent A2: 25 mM Tris-HCl pH 8.0, 40 mM NaCl, 10 mM imidazole pH 8.0, 0.5 mM ADP, 1 mM DTT). Due to the cleavage of the 6His-tagged Ubiquitin tag, Nt-β-Rca eluted in the early fractions and was subsequently pooled together. This pool was then applied to a HiTrap Q HP 5mL anion exchange column (GE) and

washed with 40 mM NaCl (Solvent C: 25 mM Tris-HCl pH 8.0, 40 mM NaCl, 0.5 mM ADP, 1 mM DTT) to remove unbound material. A step gradient to 160 mM NaCl (Solvent D: 25 mM Tris-HCl pH 8.0, 1 M NaCl, 0.5 mM ADP, 1 mM DTT) was then used to wash off weakly bound species. Desired Nt-β-Rca was eluted with a linear gradient from 160 mM to 500 mM NaCl (Rca eluted at 250 mM NaCl). All fractions containing Nt-β-Rca were pooled and concentrated down to 2.5 mL using a Centriprep concentrator (Millipore, Milford, MA). Samples were then buffer exchanged into 50 mM HEPES-NaOH, pH 7.5, 250 mM KCl, 0.5 mM ADP, 5 mM $MgCl_2$, and 10% glycerol using PD-10 columns (GE Healthcare) and subsequently concentrated back down to 1 mL. The sample was then aliquoted, flash frozen in liquid nitrogen, and stored at -80 ˚C. Protein concentrations were determined using the Bradford method (n = 9) using BSA as a standard, with typical yields of 3-5 mg/L of cell culture. Fractions from each of the various purification steps, as well as the final sample, were analyzed by SDS-PAGE to determine purity, which ranged from 85 – 98% (determined visually).

Dye conjugation of Nt-β-Rca. The labeling strategy for both "Wild-type" and the Assembly mutant involved targeting the engineered cysteine residue at position 379 of the C-terminal tail with Alexafluor dyes functionalized with a maleimide group on the 6-carbon linker. There are 5 total cysteine residues in the both constructs, however, aside from the residue on the tail, all are buried within the protein fold, and are thus not accessible for dye labeling. However, it is worth noting that when doing the labeling reaction in the presence of ADP, the protein tends to have higher dye to protein labeling ratios, suggesting that the buried cysteines may be getting exposed and thus labeled. In contrast, when labeling in the presence of ATP-γS, such over-labeling is not observed.

In brief, generating Alexa Fluor 546 C5-maleimide ($\lambda_{max\text{-}ex}$ = 554 nm; $\lambda_{max\text{-}em}$ = 570 nm; MW = 1034 Da) (Molecular Probes, Inc.) labeled Nt-β-Rca involves mixing both dye and Rca at a molar ratio of 2 to 1 (dye to protein) in a reaction mix containing 50 mM HEPES pH 7.2, 125 mM KCl, 10 mM $MgCl_2$, 5 mM ATP-γS. The total protein concentration was kept at 100 μM along with a dye concentration of 200 μM, in a total volume of 300 μL. This dye to protein molar ratio was reduced from 3.5 to 1, due to significant over labeling, even in the presence of ATP-γS. The reaction was allowed to proceed for 4 hours at 4 ˚C. This reaction time is shorter than the previously established protocol, which allowed the reaction to proceed overnight (about 14 – 16 hours). The reduction in reaction time was a result of monitoring dye reactivity as a function of time, and observing that nearly all free dye reacts with the protein within 1 hour of incubation, therefore making overnight incubation unnecessary. To remove excess free dye, 300 μL of saturated ammonium sulfate was added incrementally to the sample and incubated at 4 ˚C for an additional 30 minutes. The sample was pelleted down by micro-centrifugation in an Eppendorf tube, re-suspended in 180 μL of a buffer containing 50mM HEPES pH 7.5, 250mM KCl, 10% glycerol, 5mM $MgCl_2$, and 0.5 mM ADP, and desalted over 2 gel filtration spin columns pre-equilibrated with the same buffer (Sephadex G50-fine, 2 mL bed volume, centrifugation at 400g for 2 min). Desalted sample was then aliquoted, flash frozen in liquid nitrogen, and stored at -80 ˚C. This same method was used to generate Alexa Fluor 647 C2-maleimide ($\lambda_{max\text{-}ex}$ = 651 nm; $\lambda_{max\text{-}em}$ = 671 nm; MW = 1250 Da) (Molecular Probes, Inc.) labeled Nt-β-Rca as well as Tetramethylrhodamine (TMR) (C6

($\lambda_{max\text{-}ex}$ = 552 nm; $\lambda_{max\text{-}em}$ = 581 nm; MW = 482 Da) (Molecular Probes, Inc.) labeled Nt-β-Rca.

Rca Characterization by High-Performance Liquid Chromatography (HPLC) and Matrix Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry: Both label and label-free samples of both constructs were analyzed by reverse-phase HPLC (Agilent Technologies 1260 Infinity Quaternary LC system, Agilent 1260 Series Diode-Array detector) on a C18 analytical column (Vydac, 218TP, 5 μm, 250 mm × 4.6 mm) using a linear water/acetonitrile gradient with 0.1% trifluoroacetic acid. The absorbance of the sample was monitored at OD_{220}, OD_{280}, OD_{554} for Alexa Fluor 546, and OD_{650} for Alexa Fluor 647, to ensure that both protein and dye would be detected. The chromatograms of both the labeled and label-free "Wild-type" samples contained a single peak (elution at 46.7% B for label-free sample, elution at 48.1% B for Alexa Fluor 546 labeled sample, elution at 46.9% B for Alexa Fluor 647 labeled sample), but more importantly, lacked any free dye peaks (Elution at 42.4% B for Free Alexa Fluor 546 and Elution at 16.5% B for Free Alexa Fluor 647) suggesting that all dye present was covalently attached to the protein. The results were similar for both labeled and label-free Assembly mutant samples (elution at 49.9% B for label-free sample, elution at 51.2% B for Alexa Fluor 546 labeled sample, elution at 50.2% for Alexa Fluor 647 labeled sample).

Labeling efficiencies were determined from the absorbance of each species, using extinction coefficients (for Alexa Fluor 546, ε_{554} = 104,000 M^{-1} cm^{-1}; for Alexa Fluor 647, ε_{650} = 239,000 M^{-1} cm^{-1};) and methodology outlined by the manufacturer (Molecular Probes, Inc.). In brief, the "Wild-type" Rca concentration was determined using the absorbance at OD_{280} and the calculated extinction coefficient (ε_{280} = 39,880 M^{-1} cm^{-1}). In

order to correct for Alexa Fluor 546 dye contribution at OD_{280}, OD_{554} was multiplied by a correction factor of 0.12 and subtracted from the OD_{280} signal. Similarly, for Alexa Fluor 647, OD_{650} was multiplied by a correction factor of 0.03 and subtracted from the OD_{280} signal. The same calculations and corrections were done for the labeled Assembly mutant sample (ε_{280} = 39,880 M^{-1} cm^{-1}).

In addition to HPLC, labeled and label-free samples were also analyzed by MALDI mass spectrometry (Applied Biosystems/MDS SCIEX 4800 MALDI TOF/TOF analyzer) using sinapinic acid as the matrix. The experimentally determined mass for label-free "Wild-type" Rca was 42,786 Da, while that of Alexa Fluor 546 and Alexa Fluor 647 labeled Rca was 43,797 Da and 43,845 Da respectively. The theoretical mass of label-free "Wild-type" Rca is 42,764 Da, with the theoretical mass of Alexa Fluor 546 and Alexa Fluor 647 was 1,034 Da and 1,250 Da respectively. When working with the Assembly mutant, the experimentally determined mass for label-free Rca was 42,594 Da, while that of Alexa Fluor 546 and Alexa Fluor 647 labeled Rca was 43,637 Da and 43,897 Da respectively. The theoretical mass of label-free Assembly mutant is 42,707 Da. In both cases, the error estimate appears to be close to 100 Da.

To determine whether or not the engineered cysteine on the C-terminal tail was being specifically targeted, labeling experiments were conducted on true wild-type Nt-β-Rca (theoretical mass 42,748 Da) and "Wild-type" Nt-β-Rca, side by side using the established labeling protocol. When analyzed by reverse-phase HPLC, wild-type Nt-Rca eluted as a single peak, but did not carry the Alexa Fluor dye label, as observed for both of the other constructs. Analysis by MALDI mass spectrometry indicated that the Alexa Fluor dye label was not covalently attaching to native cysteine residues buried in the protein fold,

given that the experimentally determined mass of the labeled sample (MW 42,852 Da) was close to that of the unlabeled species (MW 42,782 Da), but rather, suggested that only the engineered cysteine on the C-terminal tail was being targeted for labeling.

FCS Experiments and Data Analysis. FCS measurements were carried out using a home-built confocal optical setup.(18) Excitation was achieved with a Compass 215M-10 532 nm CW laser (Coherent GmbH, Germany) attenuated to 60 μW to minimize triplet dynamics. The output of the laser was expanded, collimated, and directed via a dichroic lens into an Olympus PlanApo 100X/1.4NA Oil objective. Samples were placed into perfusion chambers pre-treated with BSA to minimize Rca adsorption onto the cover glass (see below). Fluorescence from samples was collected via the same objective, separated from excitation light through the dichroic, and reflected into a 50 μm pinhole. The emission was then focused into an avalanche photodiode detector (Perkin-Elmer Optoelectronics, SPCM-AQR14). A bandpass filter was used before the detector to minimize stray light (Omega 3RD560-620). The recorded fluorescence signal was autocorrelated in real time using an ALV5000/EPP USB-25 correlator (ALV GmbH, Germany).

Glass slides (Fisherbrand) were first treated in an ozonator chamber (BOEKEL UV Clean 135500) for 20 min (10 min per side), then sonicated for 45 minutes in 3% Hellmanex (Hellma) cleaning solution, washed extensively with nano-pure water and dried under N_2 gas. After placing a new glass slide with a new perfusion chamber (CoverWell) onto the objective and determining the back reflection, the perfusion chamber was treated with BSA (0.5 mg/mL) for 20 minutes, then removed and allowed to dry for an additional 20 minutes. The laser power was subsequently reduced to ~60 μW. The confocal

parameters of the observation volume were determined by measuring a 20 nM TAMRA (carboxylic acid of tetramethyl rhodamine) free dye solution ($\lambda_{max\text{-}ex}$ = 548 nm; $\lambda_{max\text{-}em}$ = 574 nm; MW = 430 Da) (Molecular Probes, Inc.) with a known diffusion coefficient (420 $\mu m^2 s^{-1}$). The average decay was fitted using the autocorrelation function to extract the values of r_0 and z_0 (Eq.1), where G is the amplitude, r_0 and z_0 are the radial and axial semiaxes of the Gaussian confocal volume, D is the diffusion coefficient and τ is the diffusion time of the diffusing species. We note that $z_0 >> r_0$ in our setup, so Eq. 1 is only marginally sensitive to the value of z_0. Typical r_0 values used in this work ranged from 0.41 µm to 0.45 µm, however most of the measurements had r_0 values closer to 0.42 µm. If the r_0 value was too large or too small, minor adjustments were made to the objective lens to either increase or decrease the back reflection (increasing the size of the back reflection gives larger r_0 values while decreasing the size gives smaller values). The chamber was rinsed with nano-pure water to remove free dye until the recorded signal matched that of pure buffer (background). For FCS data collection, 20 autocorrelation traces were collected for each sample (30 s each). Therefore, the time period for data acquisition was 10 min. Average photon count rates were about 50-90 kHz.

$$G(\tau) = G_0 \left(1 + \frac{4D\tau}{r_0^2}\right)^{-1} \left(1 + \frac{4D\tau}{z_0^2}\right)^{-\frac{1}{2}} \qquad \text{(Equation 1)}$$

Once instrumental preparations were complete, the label and label-free samples of Nt-β-Rca were removed from the -80 ˚C freezer and thawed on ice for 10 minutes. For the labeled sample, a 1 µM working solution was generated by mixing an appropriate amount

of concentrated labeled stock solution and buffer (50 mM HEPES-NaOH, pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol, and 2 mM of either ATP-γS, ADP, ATP, ADP(AlF_x), a mixture of nucleotides [80% ATP-γS and 20% ADP], or Apo) containing the particular nucleotide for the condition being tested. A working solution (either 1 μM, 10 μM, 100 μM, or no dilution) of the label-free sample was also generated using the concentrated label-free stock and the same buffer used for the labeled sample, however, each working solution was made fresh at the time of use and was kept on ice for no more than 2 hours, to minimize incubation in a particular condition. Also with each new working solution of the label-free sample, a fresh working solution of the labeled sample was also generated. Specific concentrations were sampled by mixing both label and label-free sample, keeping the labeled component fixed at 0.05 μM, and incubating on ice for 10 minutes, a time determined to be sufficient for randomly distributing the labeled subunits among all oligomers. For example, measuring a 1 μM sample involved mixing 0.05 μM labeled sample and 0.95 μM label-free sample together making a 1 μM sample. The concentration range tested was 0.05 μM to 60 μM. After generating the appropriate concentration, 35 μL of the sample was placed inside the perfusion chamber and measured 20 times, each 30 seconds, for a total acquisition time of 10 minutes. This procedure was then repeated for all concentrations in all conditions. Targeting the ADP(AlF_x) condition involved changing the methodology significantly. Instead of diluting in buffer containing the nucleotide, the sample was diluted in an Apo buffer (50 mM HEPES-NaOH, pH 7.5, 200 mM KCl, 5 mM $MgCl_2$, and 10% glycerol), 2 mM ADP and 50 mM NaF was added, and allowed to equilibrate at room temperature for 10 minutes. This was then followed by addition of 2 mM $AlCl_3$, along with another 10-minute incubation, and then measurement on the FCS

instrument. This was an adaptation form a previous group's work with this condition [24]. In an attempt to prevent systematic errors in the data collection process, protein concentrations were measured from low to high, high to low, and middle to high to low,

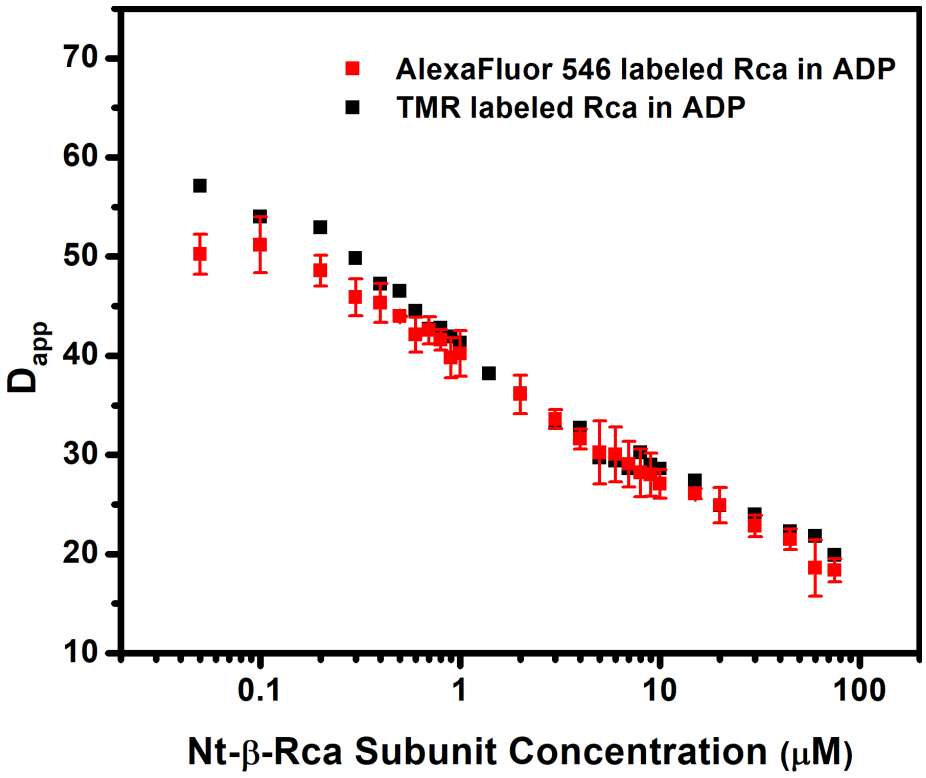


Figure 2.1: Assembly Comparison of "Wild-type" tobacco Rca labeled with different dyes. Assembly comparison between "Wild-type" Rca labeled with either AlexaFluor 546 (red data) or TMR (black data) over the same concentration range. Data for the TMR labeled sample was collected once while data for AlexaFluor labeled sample was collected in triplicate.

ensuring that each concentration experienced a different incubation time in each particular condition. All concentrations were collected in triplicate. Additionally, a control assembly experiment was conducted using material labeled with TMR and analyzed in the presence of ADP to ensure that the attachment of Alexafluor 546 was not affecting the self-assembly of Nt-β-Rca (**Fig. 2.1**).

The raw autocorrelation curves for each concentration were analyzed individually to identify and remove any traces containing aggregated material. If more than 25% of the traces were removed, the particular concentration was collected again. The remaining traces were then averaged and fitted with the autocorrelation function in equation 1 to extract apparent diffusion coefficients. The confocal parameters used for fitting were those extracted from fitting the free dye at the beginning of the experiment. Because data collection was conducted over several months, r values were determined independently for each experiment. All fits were done manually and inspected visually.

In order to extract meaningful size information from the FCS experiments, all data was normalized with respect to the D_{app} value measured at 0.05 μM, ensuring to correct for some dimeric component present at that concentration. The dimer correction was done by using experimentally determined K_d values for the "Wild-type" monomer-dimer species by collaborators in the Moerner Lab at Stanford University, in either ATP-γS, ADP, or Apo [21]. For each condition at 0.05 μM, a theoretical "pure" monomer D_{app} was calculated and used to normalize all other concentrations. This normalization method had to be used on each condition individually due to slight differences in D_{app} at the low μM concentration range. Despite only determining K_d values for the "Wild-type" sample, both the "Wild-type" and Assembly mutant samples were normalized using the same K_d values for identical condition. The K_d value used for the ATP-γS condition was 1.2 μM, while the value for the ADP was 0.5 μM. For the "Wild-type" sample in the Apo condition, the K_d value used was 0.7 μM, however, the Assembly mutant sample appeared to be monomeric for a wide concentration range, and was therefore normalized by used the average D_{app} value over the first 5 concentrations tested. Because the assembly was similar to ATP-γS

in the low μM concentration range, the ATP condition was normalized using the same K_d value of 1.2 μM. In the condition containing 80% ATP-γS and 20% ADP, the Kd value was 1.2 μM, due to the high abundance of ATP-γS. Lastly, because work with Spinach showed similar effects on exchange rates between ATP-γS and ADP(AlF_x), the ADP(AlF_x) condition for both "Wild-type" and Assembly mutant was normalized using the ATP-γS K_d value of 1.2 μM.

Monomer Characterization: Two experiments were conducted to confirm the presence of a monomeric species at low μM concentrations, as observed by collaborators in the Moerner lab. The first experiment involved measuring the lifetime of the Alexa Fluor 546 and TMR used to label the protein, both free in solution as well as attached to Nt-β-Rca, using Time Correlated Single Photon Counting (TCSPC) instrument. Unlike the FCS work, this experiment used entirely labeled material in the buffer (50 mM HEPES-NaOH, pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol, and 2 mM of nucleotide) containing either ATP-γS, ADP, or Apo, at a concentration of either 0.1 μM, 0.5 μM or 1 μM. Raw decays was analyzed using the ASUFit Software and lifetime values were extracted. Raw curves of both free Alexa Fluor 546 and TMR were fit with a single exponential and extracted lifetime values were consistent with published results. In contrast, raw data of material labeled with either dye could only be fit with three exponentials, indicating the presence of a quenched component which seemed consistent with a dimeric species. Increasing concentration lead to an increase in the quenched component, suggesting a greater degree of dimer population.

The second experiment used involved three distinct parts and used TMR labeled Nt-β-Rca in buffer (50 mM HEPES-NaOH, pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 10%

glycerol, and 2 mM ADP). The first part involved using the FCS instrument to count the total number of molecules present in the observation volume over a small concentration range (0.05 μM to 0.3 μM). This was done by using the inverse of the amplitude extracted from fitting the autocorrelation curve for each concentration. The total number of molecules increased linearly with increasing concentration, suggesting predominately monomeric species (**Fig. 2.2**). The second part involved measuring the brightness per molecule of both free TMR(ε_{FD}) and TMR labeled sample (ε). This was done by taking the average counts rate (kHz) for each of the 20 traces for a particular concentration, and dividing by the total number of molecule observed. The $\varepsilon/\varepsilon_{FD}$ was determined to be ~0.8 across the concentration range, consistent with about 1 dye per oligomer (**Fig. 2.3**). The last part involved determining the D_{app} value at each concentration tested, using the method described above. D_{app} values were determined to be constant across the concentration range (**Fig. 2.4**). The combination of all three experiments indicate the presence of a significant monomeric population in the low μM concentration regime, a result consistent with collaborators from Stanford.

Modeling Assembly and Extracting K_d values: After proper normalization, the next step was to visually fit the data with a particular assembly model. Work done on Cotton Rca suggested that the best assembly model was 1-2-4-6-24 as opposed to 1-6-24 or 1-3-6-24. However, after our previous publication, we developed a new algorithm for modeling assembly using a 1-2-4-6-12-18-…-6n where *n* is allowed to grow indefinitely as required by the data and K_d 1 = monomer-dimer, K_d 2 = dimer-tetramer, K_d 3 = tetramer-hexamer, K_d 4 = hexamer-twelvemer, and K_d 5 = twelvemer-18-…-6n, where n is an additional hexameric ring. This new algorithm thus models all assemblies larger than

twelvemers, in increments of 6, up to 132mer. The term "visually fit" was used because the process involved manually adjusting K_d values up and down to determine the best visual fit. Because the experimentally determined K_d 1 value was used for normalization,

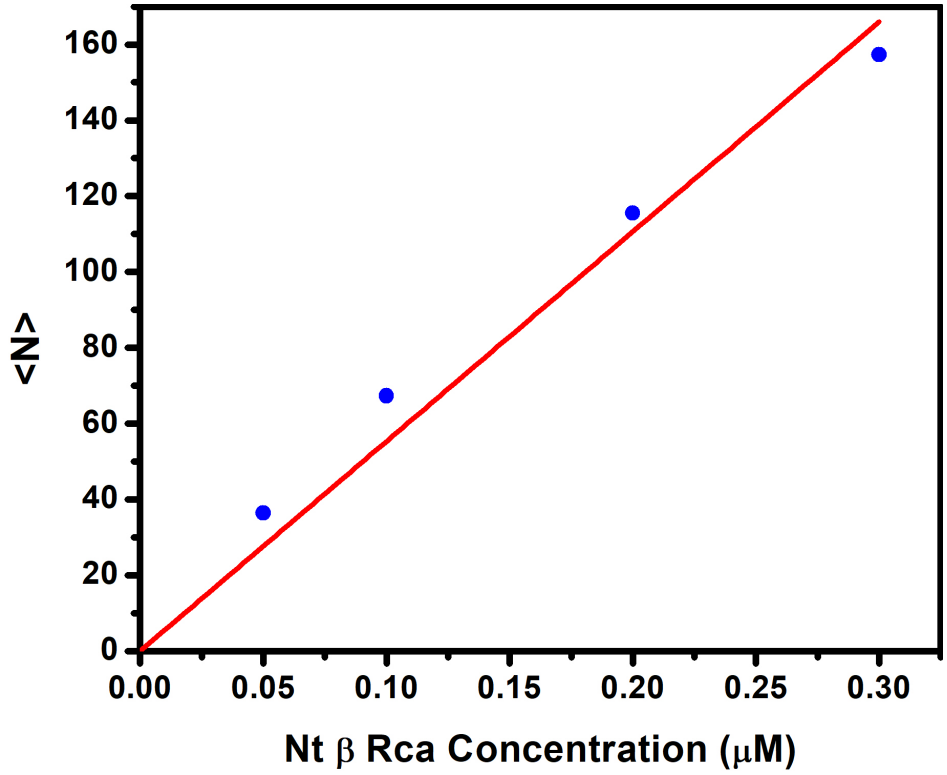


Figure 2.2: Number of Rca molecules present in the observation volume. Determination of the average number of TMR-labeled Rca molecules present in the observation volume at 4 different concentrations (blue data). With increasing concentration, there is almost linear increase (linear fit shown by red line) in the number molecules present suggesting little to no changes in overall assembly. Data for the TMR labeled sample was the average of measurements collected in triplicate.

it was kept constant throughout the fitting process. The next step was to determine the relative magnitudes of K_d 2 and K_d 3, while keeping K_d 4 and K_d 5 constant at 1. When fitting K_d 2 and K_d 3, it was found that they were highly correlated, which meant that they could vary considerably, as long as their product was kept within a certain range. Therefore, in order to properly determine the range on the product, K_d 2 was arbitrarily fixed and K_d 3 was allowed to vary. Upper and lower limits of K_d 3 were determined visually, yielding

upper and lower product limits. The process was then repeated by fixing K_d 3 and allowing K_d 2 to vary, yielding upper and lower limits of both the K_d 2 value and the product range.

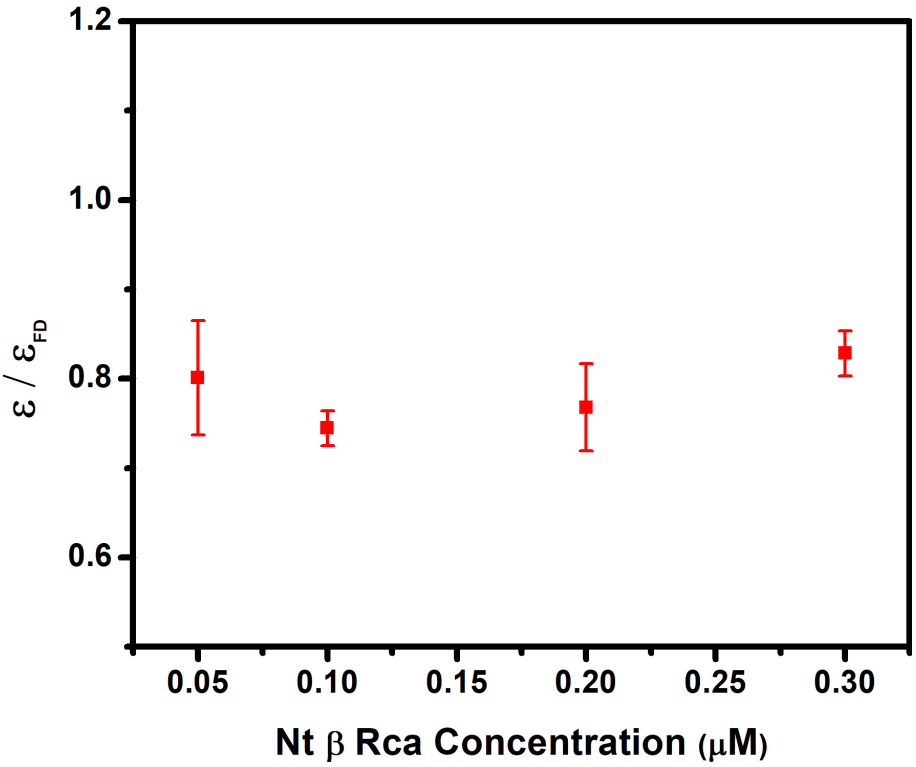


Figure 2.3: Brightness comparison between Free TMR dye and TMR-labeled Rca. Comparison between the brightness of "Wild-type" Rca labeled with TMR normalized with respect to the brightness of the free TMR dye. These brightness values are calculated per molecule of either Rca or free dye, rather than just the overall brightness. Data for the both TMR labeled sample and free TMR dye were collected in triplicate.

The upper and lower, as well as the average, limits were then fixed and both K_d 2 and K_d 3 were both allowed to vary, trying to identify which K_d values resulted in the maximum tetramer component. With K_d 2 and K_d 3 determined, they were then kept constant and K_d 4 and K_d 5 were allowed to vary. It is worth noting that K_d 4 and K_d 5 were not correlated with each other, nor are they correlated with K_d 2 and K_d 3. Like with K_d 2 and K_d 3, upper and lower limits were determined visually for K_d 4 and K_d 5. This procedure was repeated for both the "Wild-type" and Assembly mutant samples in all conditions. The data

presented shows the fits which yielded average tetrameric species.

Thermofluor Assays: A Thermofluor assay was used to extract apparent "melting"

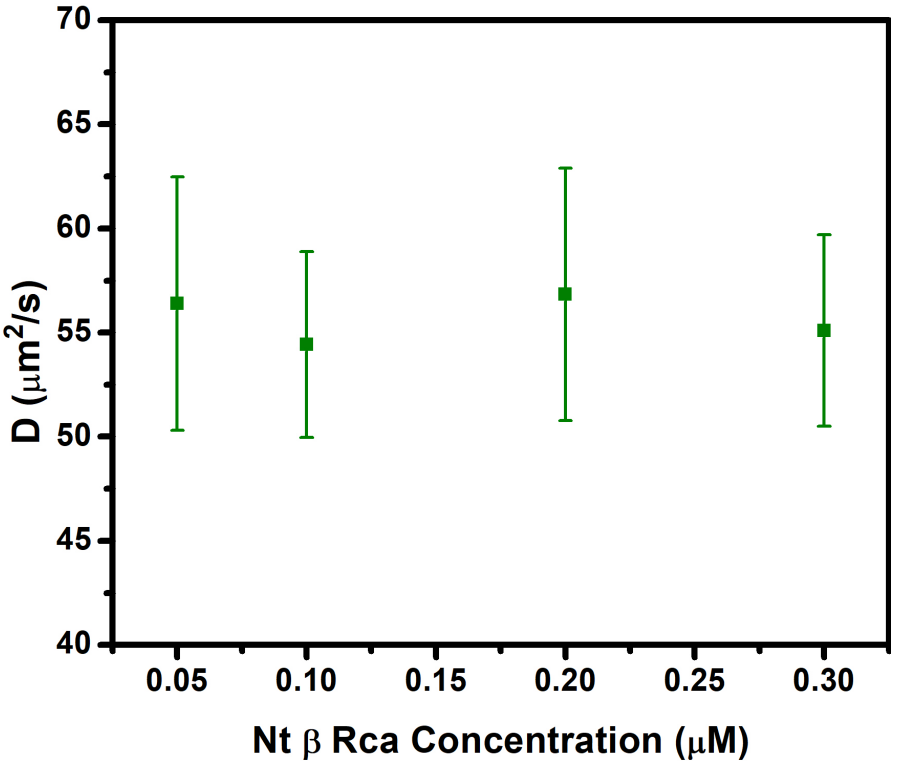


Figure 2.4: Comparison of D_{app} values from TMR-labeled Rca over the low μM range. Comparison between D_{app} values determined for "Wild-type" Rca labeled with TMR dye. Over the concentration range tested (0.05 to 0.3 μM) the D_{app} values remained relatively constant, suggesting little to no changes in assembly. Data for the TMR labeled sample was collected in triplicate.

temperatures (T_m) as a means of better examining the thermal-stability of both Nt-β-Rca constructs in various nucleotides. This was achieved by generating a premix containing 0.25 mg/mL Nt-β-Rca (either "Wild-type" or Assembly Mutant), 50 mM HEPES pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 10% Glycerol, and 16X SYPRO Orange (Invitrogen, Molecular Probes). 18 μL of premix was then added to each well of a 384-well polypropylene TempPlate PCR plate as well as 2 μL of concentrated nucleotide stock

(ATP-γS, ADP, ATP, ADP(AlF_x), or 2 µL of H_2O for the Apo condition) for a total volume of 20 µL and a total protein concentration of 0.23 mg/mL. Prior to placement in thermal cycling instrument (Applied Biosystems ABI Prism 7900HT Sequence Detection System), the plate was covered with an optically clear seal and centrifuged for 1 minute, to ensure all contents were properly mixed. The method used for each runs consisted of three stages: 1) a fast ramp (~0.5 °C/s) from 25 °C to 4 °C, where the temperature was held for two

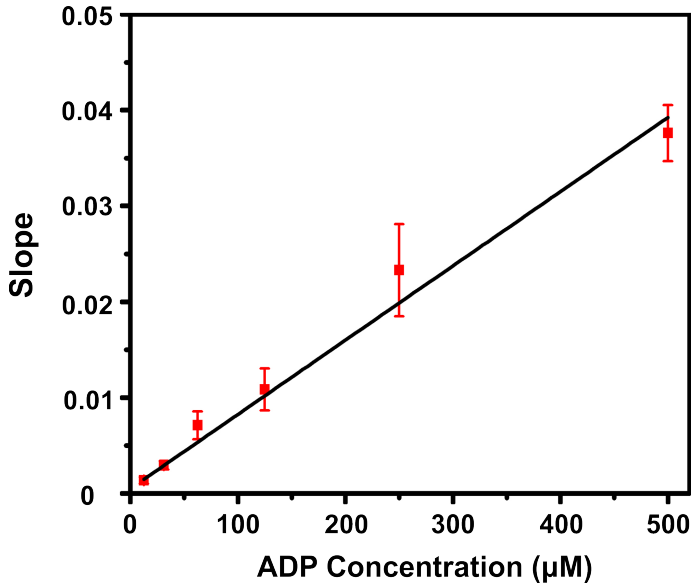


Figure 2.5: ADP standard curve for NADH assay. To ensure that NADH oxidation was not the rate limiting step of the assay, various ADP concentrations were tested to ensure a linear response from the assay without Rca. At concentrations higher than 500 µM ADP, the assay response is no longer linear. Slopes measured with Rca fell within the linear response range of the assay. (Data points need to be repeated to determine error bars)

minutes; 2) a slow ramp (~0.03 °C/s) from 4 °C to 80 °C holding for two minutes; 3) a fast ramp (~0.5 °C/s) down to 4 °C, with fluorescence data being collected every ~8.5 seconds for the entire collection time. The apparent melting temperature (T_m) was approximated by using the equation shown below:

$$I_{Tm} = B - \frac{1}{2}(B - A) \qquad \textbf{Equation 2}$$

where A is the minimum intensity prior to the fluorescence increase and B is the maximum in intensity following the increase.

Continuous ATPase Assays Based on ADP Production (NADH Assay): In order to monitor the ATPase activity of Nt-β-Rca as a function of concentration, an enzyme-coupled assay which monitors NADH oxidation as a result of ADP production was implemented, as described previously [17]. The working solution for the assay contained 100 mM Tricine (pH 8.0), 20 mM KCl, 5 mM DTT, 5 mM $MgCl_2$, 2 mM phosphoenolpyruvate, 0.3 mM NADH, 1.3−2.2 units of pyruvate kinase, and 2−3.1 units of lactate dehydrogenase (Sigma-Aldrich). In order to test the concentration range yielding a linear response, an ADP standard curve was generated by mixing 25 µL of various known ADP concentration, with 375 µL of assay mix (400 µL total volume). These ADP concentrations, ranging from 12.5 µM to 500 µM, were made by combining an appropriate amount of ADP stock and buffer (50mM HEPES pH 7.5, 5mM $MgCl_2$, 250mM KCl, 10% glycerol). All samples were monitored at 340 nm for 40 seconds prior to addition of ADP, to ensure a stable signal was achieved, and then for an additional 260 seconds after addition of ADP, yielding a total measuring time of 300 seconds. NADH oxidation rates were then calculated based on linear fits of the chromatograms which gave highest rates over a 36 second period, using a known extinction coefficient of 6220 $M^{-1}cm^{-1}$. Based on the standard curve, ADP concentrations up to 500 µM showed a highly linear response (**Fig. 2.5**).

With the assay's linear response established, various concentrations of "Wild-type" tobacco Rca were tested to examine their ATPase activity. This was done by mixing

various concentrations of Rca with the same buffer described above, using a dilution scheme similar to one used in FCS experiments to ensure proper comparison. 25 µL of the Rca protein solution was then mixed with 373 µL of assay mix, with final Rca concentrations ranging from 0.25 µM to 20 µM, and measured for ~40 seconds, to ensure a stable signal was achieved. The sample was then spiked with 2 mM ATP and measured for an additional ~460 seconds, yielding a total measuring time of 500 seconds. NADH oxidation rates were determined as before, using a linear fit to determine the steepest rate over a 36 second period. Due to a 1 to 1 stoichiometric ratio between NADH oxidation and ADP production, the rate of NADH oxidation was equal and opposite to the rate of ADP formation, thus allowing us to determine the rate of ADP production per subunit of Rca, per minute, with rates being measured a minimum of 5 times. This same methodology was repeated for ATPase activity characterization of the Assembly Mutant construct.

End Point ATPase Assays Based on ADP Production (NADH Assay): Unlike the continuous assay, the endpoint assay was utilized to measure the total amount of ADP produced over the course of 30 minutes in order to quantify how much ADP was generated in the Active ATP Turnover condition tested by FCS. Because ATPase activity is concentration dependent, 3 different concentrations were tested, 1.5 µM, 5 µM, and 15 µM, given that this region gave the most variation in oligomer distribution. Each Rca concentration was generated by mixing an appropriate amount of stock solution with a buffer containing ATP (50mM HEPES pH 7.5, 5mM $MgCl_2$, 250mM KCl, 10% glycerol, and 2 mM ATP) and incubated at room temp. Immediately after mixing with the buffer, 20 µL was taken and added to a quartz cuvette (VWR International) containing 380 µL of the working solution (100 mM Tricine (pH 8.0), 20 mM KCl, 5 mM DTT, 5 mM $MgCl_2$, 2

mM phosphoenolpyruvate, 0.3 mM NADH, 1.3−2.2 units of pyruvate kinase, and 2−3.1 units of lactate dehydrogenase (Sigma-Aldrich)). This sample was then monitored at 340 nm for 100 seconds. ADP production was calculated by using the change in absorbance at 340 nm from time 0 seconds to time 100 seconds. In addition to ATP, identical experiments were carried out using ATP-γS instead of ATP in order to determine whether tobacco Rca could turnover ATP-γS. These experiments were carried out for both the "Wild-type" and the Assembly mutant samples.

Results:

New Algorithm for Modeling Tobacco Assembly. In order to better understand the role of nucleotides in regulating the self-assembly pathway of Tobacco Rca, we have used previously established FCS methods to monitor the assembly of tobacco Rca from monomers to large supramolecular complexes of as much as 132 subunits, in the presence of various nucleotides [12,13]. As discussed earlier, protein preparation for FCS experiments involved purification of a recombinantly expressed fusion protein in *E. coli* containing short-form Rca and an N-terminal 6His-ubiquitin tag, by Ni-affinity and anion-exchange chromatography. Removal of the tag was carried out by incubation with a deubiquitylating enzyme (DUB). As with cotton [13], the strategy for attachment of a fluorescent labeled involved use of an engineered Cys residue on the C-Terminal tail, incorporated via an Ser-Cys substitution at the 379 position (tobacco numbering), thus providing a site for thiol directed labeling of the AlexaFluor-546 by means of maleimide chemistry. Analysis by HPLC and MADLI-TOF confirmed successful incorporation of a single dye per subunit of Rca, with molar dye to protein ratios of 1.0 ± 0.09 : 1 (n=3). Additionally, control experiments with wild-type protein lacking the engineered cysteine

residue showed poor labeling efficiency, with molar dye to protein ratios of 0.01:1, suggesting that attachment of the dye was specific to the engineered Cys residue.

Apparent (average) diffusion coefficients (D_{app}) were measured on a small number of particles, with subunit concentrations of label-free Rca ranging from 0.05 to 60 µM and

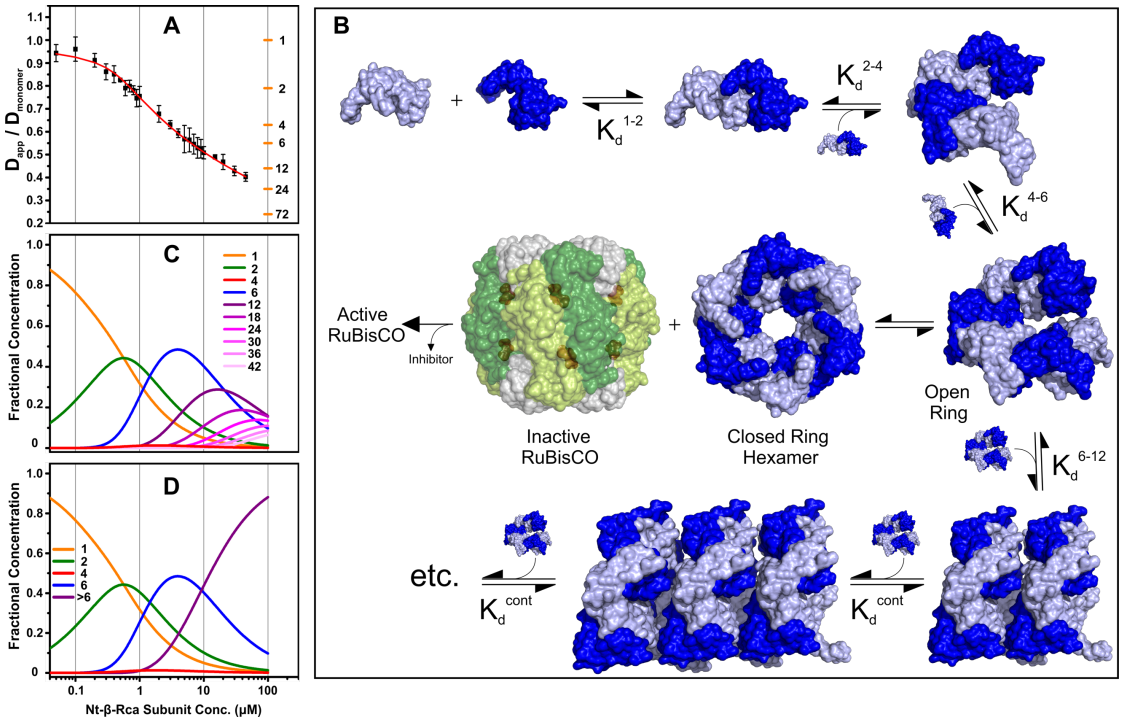


Figure 2.6: New assembly model for fitting FCS data. (A) overlay of FCS data from "Wild-type" Rca in ADP with fit of new model, shown in red. (B) New assembly model with respective K_d value for each assembly step (PDB codes; 3T15, 3ZW6, and 1RCX). (C) Fractional concentration profile of each oligomer (modeled up to 132-mer, however, shown up to 42-mer) indicated by a purple gradient where lighter colors represent larger oligomers, or (D) summed up together as one species and represented by one line.

dye-labeled Rca concentrations being kept constant at 0.05 µM, to generate an assembly curve for each nucleotide condition (**Fig. 2.6A**). The horizontal orange bars are calculated values using the Stokes-Einstein relation of the ratio of $D_{app}/D_{monomer}$ for all different assemblies, where 1, 2, 3, 4, 6, 12, 24, and 72 represents the number of subunits per

assembly. As described previously [12,13], the data were mathematically modeled in terms of individual K_d values (**Fig. 2.6A**). Attempts were made to determine K_d 1 values independently, however, a variety of different values, ranging from 0.6 to 25, were found to give equally good fits. Therefore, because K_d 1 is poorly determined by FCS, we fixed this value to the one obtained from the assembly work done using the ABEL trap set up

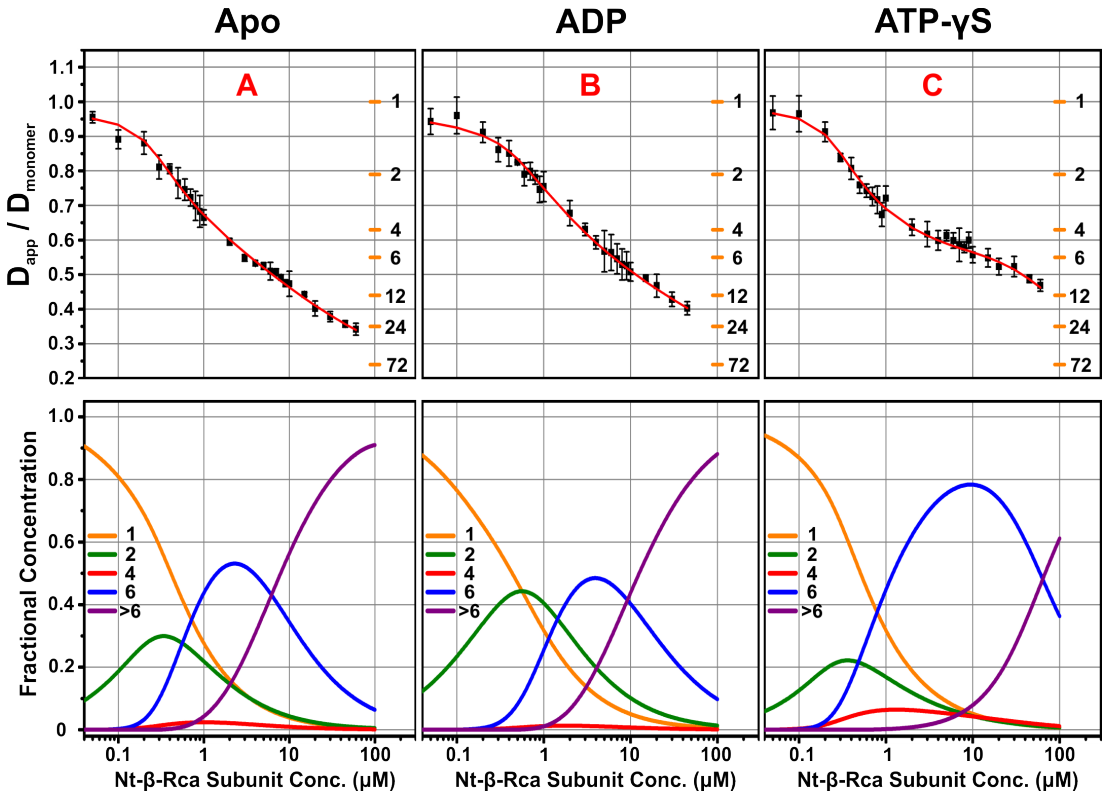


Figure 2.7: FCS Data and Assembly Profiles for "Wild-type" Rca in either Apo, ADP, or ATP-γS conditions. FCS data collected on "Wild-type" Rca in either (A) Apo, (B) ADP, (C) ATP-γS, condition and fit with new assembly model. The horizontal orange bars are calculated values using the Stokes-Einstein relation of the ratio of $D_{app}/D_{monomer}$ for all different assemblies, where 1, 2, 3, 4, 6, 12, 24, and 72 represents the number of subunits per assembly. Subsequent fractional concentration plots for each oligomer are shown in the bottom panel.

[21]. In this way, we were able to estimate K_d values for the dimer-tetramer (K_d 2), tetramer-hexamer (K_d 3), hexamer-12mer (K_d 4), and twelvemer-18-…-6n (K_d 5), where n

is the addition of a hexameric ring (**Fig. 2.6B)**. Extensive mathematical simulations of FCS data collected provided evidence that the numerical product K_d 2 × K_d 3 remains approximately constant for all K_d combinations that yield a good fit to the data. This observation suggests that K_d 2 and K_d 3 are highly correlated (table 2.1). A very large Kd

Table 2.1: Estimated equilibrium constants for tobacco Rca self-assembly

	ATP-γS			ADP			Apo			ATP			ADP(AlF$_x$)			Nuc. Ratio		
	"Wild-type"																	
K_d 1 (μM)	1.2			0.5			0.7			1.2			1.2			1.2		
K_d 2 (μM)	0.30	**0.45**	0.60	9.5	**14.0**	18.5	1.0	**2.0**	3.0	0.11	**0.13**	0.15	0.030	**0.050**	0.070	0.20	**0.29**	0.38
K_d 3 (μM)	0.040	**0.018**	0.010	0.030	**0.014**	0.0070	0.018	**0.0090**	0.0045	0.36	**0.31**	0.26	0.15	**0.11**	0.070	0.31	**0.24**	0.17
K_d 4 (μM)	32 ± 4			2.0 ± 1			1.6 ± 0.3			4.3 ± 1			0.50 ± 0.2			4.0 ± 2		
K_d 5 (μM)	10 ± 0.5			2.5 ± 0.5			1.4 ± 0.2			2.5 ± 0.2			1.0 ± 0.2			4.0 ± 0.7		
K_d 2 x K_d 3 Range	0.0050 – 0.014			0.13 – 0.22			0.014 – 0.022			0.030 – 0.045			0.0030 – 0.0081			0.044 – 0.084		
Average	0.0095			0.14			0.018			0.038			0.0055			0.064		
	Assembly Mutant																	
K_d 1 (μM)	1.2			0.5			25			1.2			1.2			-		
K_d 2 (μM)	0.24	**0.33**	0.42	3.0	**14.0**	25.0	0.50	**2.0**	3.5	0.028	**0.051**	0.080	0.050	**0.080**	0.11	-	-	-
K_d 3 (μM)	0.060	**0.045**	0.030	0.090	**0.022**	0.010	0.070	**0.050**	0.030	0.43	**0.29**	0.15	0.11	**0.080**	0.050	-	-	-
K_d 4 (μM)	65 ± 15			3.0 ± 1			2.5 ± 0.5			5.0 ± 1			65 ± 11			-		
K_d 5 (μM)	18.5 ± 1			2.0 ± 1			3.3 ± 0.3			2.5 ± 0.3			8.5 ± 0.6			-		
K_d 2 x K_d 3 Range	0.013 – 0.022			0.080 – 0.41			0.021 – 0.19			0.0035 – 0.019			0.0028 – 0.010			-		
Average	0.018			0.25			0.11			0.012			0.0064			-		

Extracted K_d values from fits of the FCS data for both "Wild-type" and the Assembly mutant in different nucleotides. The assembly model used is a 1-2-4-6-12-18-…-6n where K_d 1 is monomer-dimer, K_d 2 is dimer-tetramer, K_d 3 is tetramer-hexamer, K_d 4 is hexamer-twelvemer, and K_d 5 is 12-18-…-6n. The value for K_d 1 was fixed to a previously determined value (Wang, Q. et al) while others were determined visually. Error estimates for K_d 4 and K_d 5 were determined by visually estimating the upper and lower limits, while error estimates for K_d 2 and K_d 3 were calculated by examining the product of the two respective K_d values, given their high degree of correlation. For all condition, three sets of values are shown for K_d 2 and K_d 3, showing the lower limit, average, and upper limit. The average values bolded were used for generating the fits of the FCS data.

2 works well in combination with a very small K_d 3, providing the minimum tetramer concentration consistent with experiment (table 2.2). To the contrary, a very small K_d 2 in combination with a very large K_d 3 yields a surprisingly poor fit to the data, suggesting that

the experimental D_{app} values set a distinct limit to the amount of tetramer that may accumulate. The remaining two K_d values, K_d 4 and K_d 5 were not correlated with any other K_d thus allowing for better error estimation. Table 2.1 shows all K_d values determined for all nucleotide conditions as well as the upper and lower limits. As before, these K_d values were used to extract fractional concentration profiles showing the relative

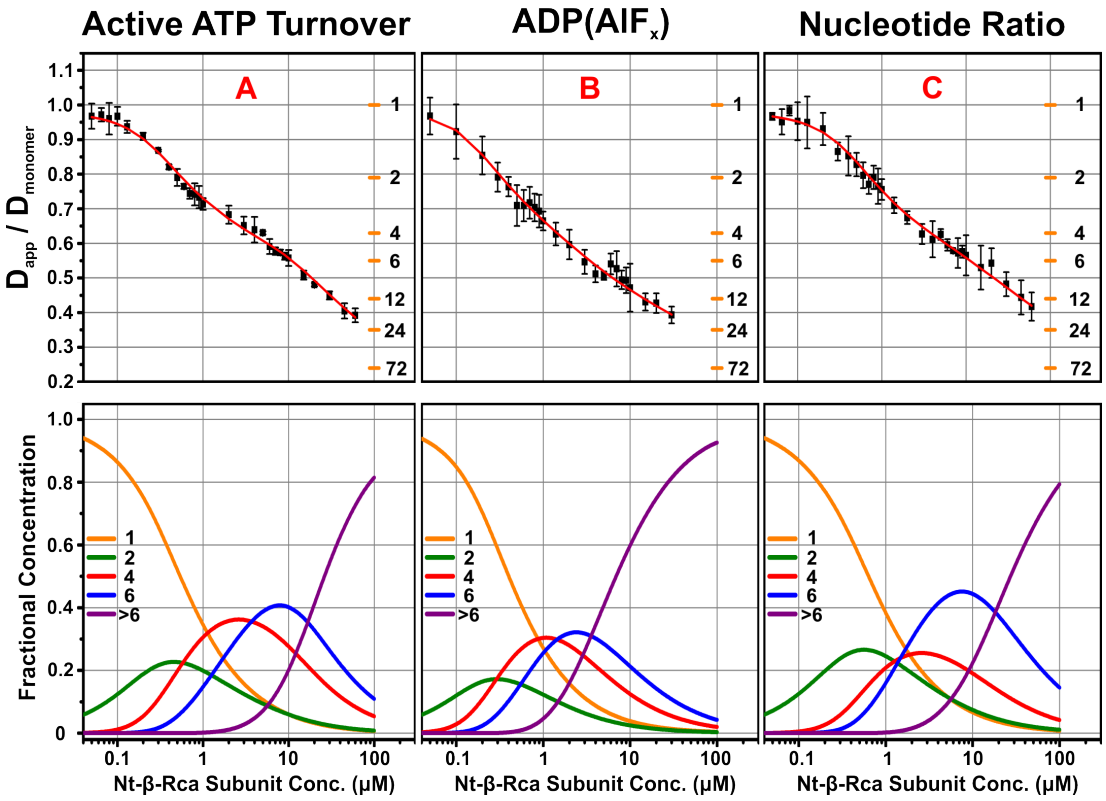


Figure 2.8: FCS Data and Assembly Profiles for "Wild-type" Rca in either Active ATP Turnover, ADP(AlF$_x$), or Nucleotide Ratio conditions. FCS data collected on "Wild-type" Rca in either (A) Active ATP Turnover, (B) ADP(AlF$_x$), (C) Nucleotide Ratio (80% ATP-γS and 20% ADP), condition and fit with new assembly model. The horizontal orange bars are calculated values using the Stokes-Einstein relation of the ratio of $D_{app}/D_{monomer}$ for all different assemblies, where 1, 2, 3, 4, 6, 12, 24, and 72 represents the number of subunits per assembly. Subsequent fractional concentration plots for each oligomer are shown in the bottom panel.

contribution of each oligomer over the concentration range tested (**Fig. 2.6C**). Despite only showing the fractional concentration of oligomers at or smaller than 42-mers, we have

identified the contribution of oligomers up to 132-mer and have summed up all fractional concentrations larger than hexamers as a single species, shown as a purple line in (**Fig. 2.6D**).

ATP hydrolysis stimulates tetramer accumulation

By using the data collection and analysis methodology discussed above, we find that for "Wild-type" Rca in the presence of either ADP or ATP-γS as the sole nucleotide, the tetramer does not accumulate to any great extent (at most 1% and 6%, respectively). In the presence of ADP, monomers, dimers and hexamers are equally populated at 1.5 μM

Table 2.2: Fractional composition of the tetramer and hexamer

	ATP-γS		ADP		Apo		ATP		ADP(AlF_x)		Nuc. Ratio	
	"Wild-type"											
	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex
Min Tet	4.1% at 1.2 μM	80.2% at 8.6 μM	0.8% at 1.6 μM	50.4% at 3.5 μM	1.4% at 0.9 μM	54.6% at 2.1 μM	32.6% at 2.4 μM	42.7% at 7.3 μM	22.7% at 1.0 μM	35.6% at 2.0 μM	20.0% at 2.4 μM	48.1% at 6.6 μM
Ave. Tet	6.4% at 1.3 μM	78.3% at 9.4 μM	1.3% at 1.8 μM	48.4% at 4.0 μM	2.4% at 1.0 μM	53.1% at 2.3 μM	36.2% at 2.6 μM	40.7% at 7.8 μM	30.4% at 1.1 μM	32.1% at 2.4 μM	25.5% at 2.6 μM	45.1% at 7.5 μM
Max Tet	11.4% at 1.5 μM	74.5% at 10.9 μM	2.3% at 2.0 μM	46.4% at 4.3 μM	4.7% at 1.1 μM	52.0% at 2.4 μM	39.9% at 2.8 μM	38.8% at 8.6 μM	39.5% at 1.2 μM	28.9% at 2.8 μM	31.8% at 1.5 μM	42.2% at 8.4 μM
	Assembly Mutant											
	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex	% Tet	% Hex
Min Tet	8.7% at 1.5 μM	80.7% at 15.8 μM	0.9% at 2.0 μM	50.7% at 4.4 μM	2.0% at 6.3 μM	35.7% at 10.1 μM	32.1% at 1.8 μM	47.4% at 6.4 μM	18.4% at 1.4 μM	75.7% at 14.7 μM	-	-
Ave. Tet	11.7% at 1.7 μM	78.7% at 17.0 μM	1.7% at 2.2 μM	49.4% at 4.6 μM	3.3% at 6.4 μM	35.4% at 10.3 μM	45.4% at 2.4 μM	39.2% at 8.9 μM	25.3% at 1.5 μM	71.2% at 17.0 μM	-	-
Max Tet	15.2% at 1.7 μM	76.7% at 18.3 μM	7.0% at 2.2 μM	47.6% at 4.8 μM	7.0% at 5.4 μM	37.0% at 9.2 μM	57.5% at 2.9 μM 4.3	32.4% at 12.2 μM	33.2% at 1.7 μM	66.6% at 19.7 μM	-	-

This table shows the percentage of tetramer and hexamer present for each nucleotide condition using the 3 different sets of Kd values listed in table 1. The concentration listed for each percentage corresponds to the concentration at which that maximum oligomer is observed.

Rca, whereas at 3.5 μM, the hexamer peaks at a fractional contribution of 48% (**Fig. 2.7B**). At yet higher concentrations, assembly proceeds in a continuous fashion to form larger and

larger aggregates (**Fig. 2.7B**) (**Table 2.1**). With ATP-γS, the monomer and hexamer appear to be equally populated at 0.8 μM (37% contribution each), exactly as observed by single molecule diffusometry [21]. The hexamer dominates over a broader concentration range, reaching a peak contribution of 78% at 9.4 μM Rca, while aggregation to larger states is substantially reduced (**Fig. 2.7C**) (**Table 2.1**). Stabilization of the hexamer by ATP-γS is in line with our reports on cotton β-Rca by FCS [13], as well as with assembly work done on tobacco Rca using ABEL trap instrument [21]. Additionally, assembly devoid of nucleotide (residual ADP concentrations ranging from 0.5 to 60 μM), a condition referred to further as Apo, proceeds much like ADP, however, at a much faster rate, with monomer, dimer, and hexamer being nearly equally populated at 0.74 μM (about 30% contribution each), and 53% maximum hexamer contribution at 2.5 μM (**Fig. 2.7A**). Like ADP, larger and larger aggregates are formed with increasing concentration. It is worth noting that K_d 2 values determined for these three conditions are in agreement with previously determined values [21], however, differ slightly due to use of average K_d values rather than either upper or lower limits, as seen in **Table 2.1**.

In contrast to ADP, ATP-γS, and Apo, addition of ATP strongly stimulates tetramer formation. Tetrameric assemblies accumulate to a maximum of 36% at 2.6 μM Rca (**Fig. 2.8A**). At 3.5 μM Rca, tetramers and hexamers are equally populated (35% contribution from each), and at 7 μM Rca, the hexamer peaks with a 41% contribution (**Fig. 2.8A**) (**Table 2.2**). These observations suggest that tetramer formation requires a mix of substrate and product nucleotides, a condition that is generated by active ATP hydrolysis during FCS measurements. Between 1.5 and 5 μM Rca, the average ATP/ADP ratio was determined to range from ~6 – 2**.** At higher Rca concentrations, the ATP/ADP ratio dropped

significantly lower to ~0.9, and in this regime, assembly to larger and larger aggregates is observed, as with ADP. This would seem to suggest that when the ATP/ADP ratio is nearly

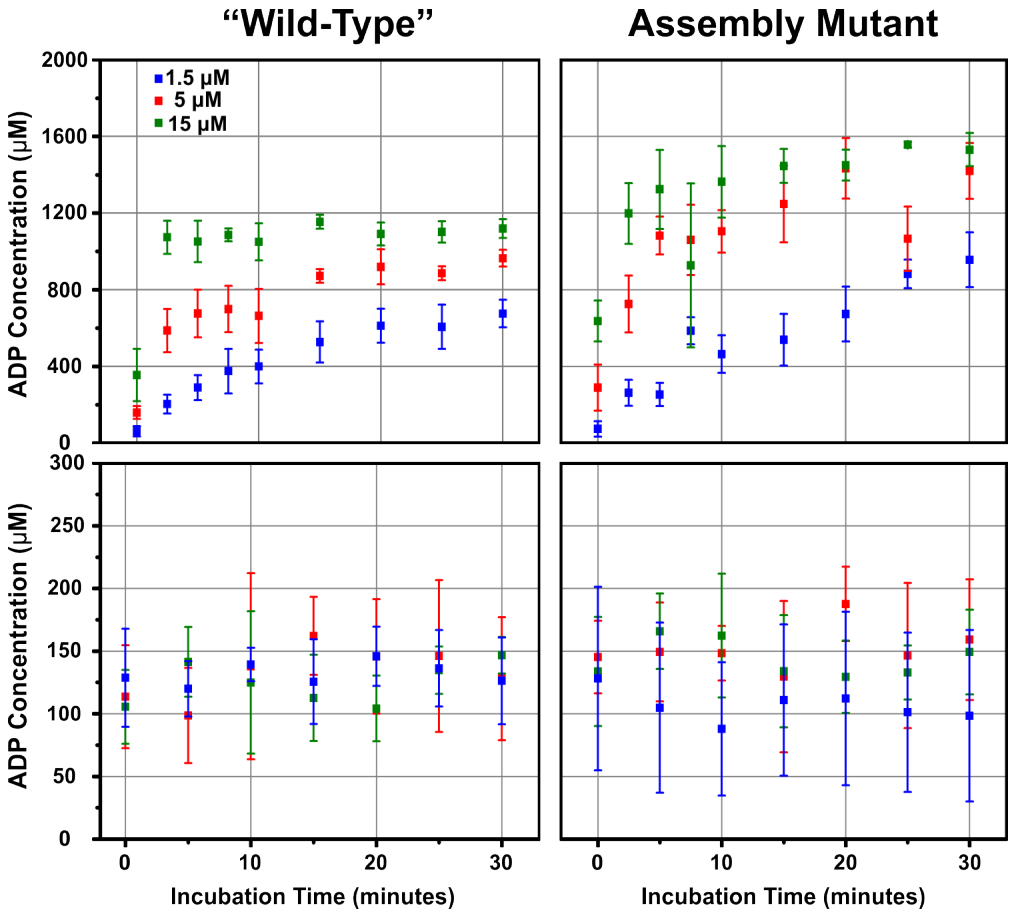


Figure 2.9: ADP accumulation in both "Wild-type" and Assembly Mutant in the presence of ATP or ATP-γS. Data collected using the NADH coupled assay monitoring production of ADP for both the "Wild-type" and Assembly Mutant samples in the presence of either ATP (top panel) or ATP-γS (bottom panel) as a function of time at three different concentrations, 1.5 μM (blue data), 5 μM (red data), and 15 μM (blue data). All samples were kept at room temperature (RT) for the entire duration of the experiment. All data shown was generated from at least 3 repeats.

equal, assembly proceeds as with ADP only and the nucleotide ratio has no effect.

Significant tetramer accumulation is also observed in the presence of ADP(AlF_x), a transition state analog generated by addition of excess ADP, NaF, and $AlCl_3$, with

maximum tetramer assemblies accumulating to 30.4% at 1.1 μM Rca (**Fig. 2.8B**) (**Table 2.2**). Thermofluor experiments in this condition show melt curves that are biphasic, suggesting site occupancy of either ADP or ADP(AlF_x), thus generating a similar mix of substrate and product nucleotides as observed in the active ATP turnover condition. Despite not knowing the exact concentration of the formed transition state analog, assembly at higher concentrations seems to promote larger and larger assemblies, suggesting ADP mediated assembly is dominant.

To determine whether tetramer accumulation was a direct result of increasing ATP/ADP ratios, a control experiment was done using a fixed ratio of nucleotides, 80% ATP-γS and 20% ADP yielding a ATP-γS/ADP ratio of 4, to determine if and how much tetramer accumulates. ATP-γS was selected over ATP due to tobacco Rca's inability to hydrolyze said nucleotide (**Fig. 2.9**). It is worth noting that ATP-γS has a 6.6 ± 0.4% ADP impurity, which was accounted for in the total ADP concentration. When compared to the ATP-γS and Active ATP turnover conditions, which have nucleotide ratios of 14 and 2 respectively (Fig. 7C and 8A), it appears that tetramer accumulation reaches 26% at 2.6 μM (**Fig. 2.8C**). This is more than the 6% tetramer accumulated in the ATP-γS condition but less than the 36% observed in the Active ATP turnover condition (**Table 2.2**). This finding seems to suggest that subtle fluctuations in nucleotide ratios seem to drastically alter, and perhaps regulate, different oligomeric distributions.

High similarity in self-assembly pathway between "Wild-type" and the Assembly Mutant. In addition to the "Wild-type" sample, we also examined an assembly mutant containing the subunit interface substitution R294V (tobacco numbering), which has proven useful in negative stain EM reconstructions, further referred to as Assembly

Mutant [11]. This mutation disrupts the hydrogen bonding between the amide group on the Arg residue at position 294 (tobacco numbering) and the carboxyl group of the Asn residue at position 99 (tobacco numbering) of the adjacent subunit. Like wild-type, assembly in the presence of ADP or ATP-γS as the sole nucleotide results in little tetramer accumulation (at most 2% and 12% respectively). In the presence of ADP, monomers, dimers and hexamers are equally populated at 1.5 μM Rca, whereas at 4.6 μM, the hexamer peaks at

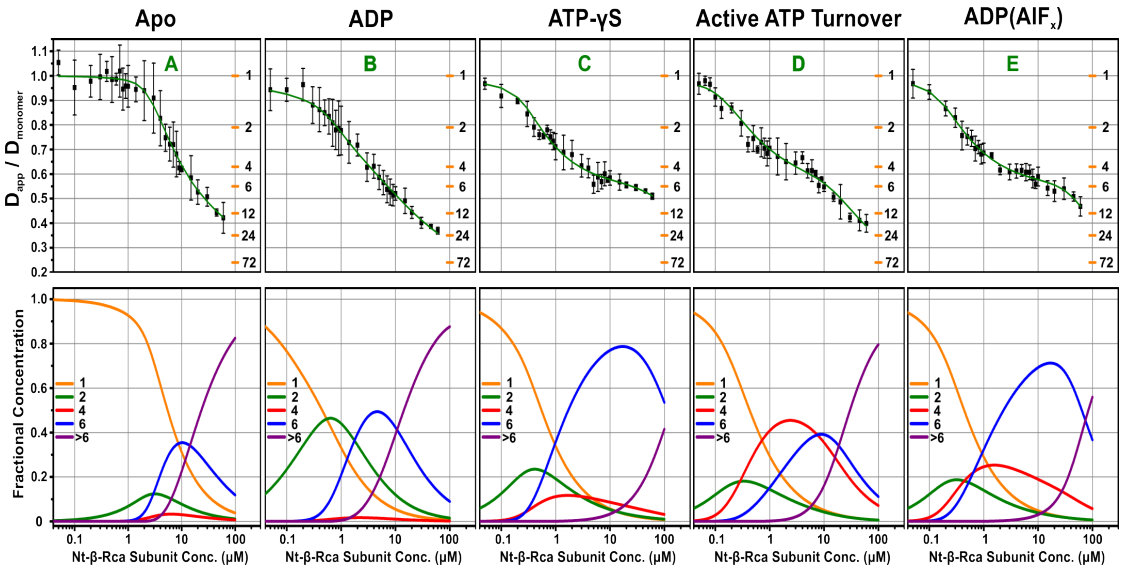


Figure 2.10: FCS Data and Assembly Profiles for Assembly Mutant Rca in all conditions. FCS data collected on Assembly Mutant Rca in either (A) Apo, (B) ADP, (C) ATP-γS, (D) Active ATP Turnover, (E) ADP(AlF_x), condition and fit with new assembly model. The horizontal orange bars are calculated values using the Stokes-Einstein relation of the ratio of $D_{app}/D_{monomer}$ for all different assemblies, where 1, 2, 3, 4, 6, 12, 24, and 72 represents the number of subunits per assembly. Subsequent fractional concentration plots for each oligomer are shown in the bottom panel.

a fractional contribution of 49% (**Fig. 2.9B**). At yet higher concentrations, assembly proceeds in a continuous fashion to form larger and larger aggregates (**Fig. 2.10B**). With ATP-γS, the monomer and hexamer appear to be equally populated at 1 μM (35% contribution each). The hexamer dominates over the same broad concentration range,

reaching a peak contribution of 79% at 17 μM Rca, with aggregation to larger states being reduced even more substantially than in "Wild-type" (**Fig. 2.10C**), thus suggesting a greater abundance of closed-ring hexamers. Unlike the "Wild-type" sample, the Assembly Mutant in the absence of nucleotide (residual ADP concentrations ranging from 0.5 to 65 μM) is actually assembly deficient, with the monomer being the only predominant species

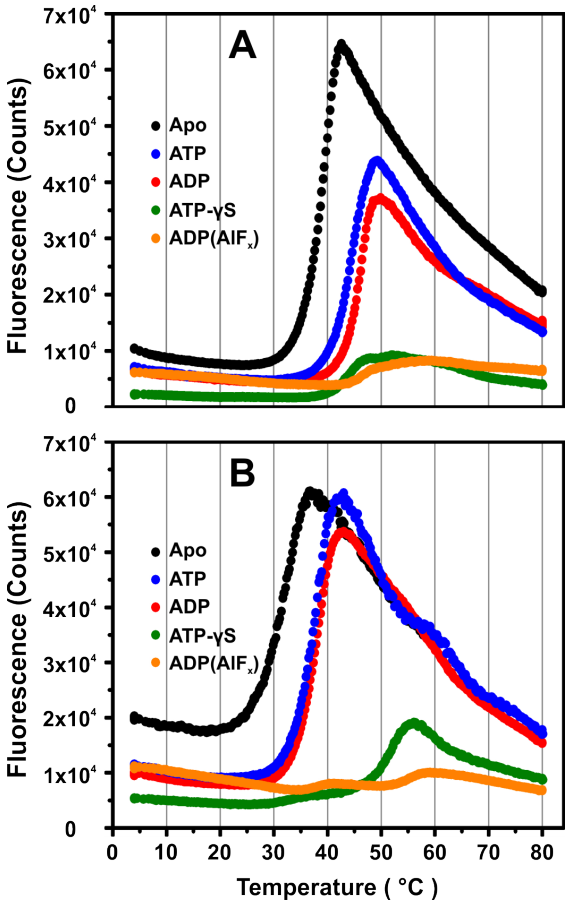


Figure 2.11: Thermal stability of both "Wild-type" and Assembly Mutant. Raw thermal melt curves for both the (A) "Wild-type" and (B) Assembly Mutant using a previously developed Thermofluor assay. Conditions testes are identical to those used in FCS, however, the protein concentration was kept constant at 5 μM. Each condition was repeated in triplicate.

up to 2 μM (**Fig. 2.10A**). At higher concentrations, the monomer and hexamer are equally populated at 9 μM (35% contribution each), but larger and larger aggregates quickly form

with increasing concentration (**Fig. 2.10A**).

As with "Wild-type", tetramer accumulation is only observed in the presence of a mixture of substrate and product nucleotides, a criterion that is only met in both the active ATP turnover and ADP(AlF_x) conditions. Addition of ATP promotes tetramer formation with a maximum of 45% at 2.4 μM (**Fig. 2.10D**). At 7.5 μM, the tetramer and hexamer are equally populated (38% contribution each), and at 8.9 μM, 39% maximum hexamer is formed, a trend quite similar to the "Wild-type" sample (**Table 2.2**). Similarly, between 1.5 and 5 μM Rca, the average ATP/ADP ratio was determined to be ~5 - 1, however, with increasing Rca concentrations, the ATP/ADP ratio dropped significantly lower to ~0.7, and in this regime, assembly to larger and larger aggregates is observed, as with ADP. At higher concentrations, ATP/ADP ratios for the "Wild-type" sample are close to 1 while this ratio for the Assembly mutant sample is closer to 0.5. This would suggest that the disruption of the hydrogen bond between Arg 294 and Asn 99 perturbs Rca's ability to turn off hydrolysis of ATP, thus generating more ADP **(Fig. 2.9)**. Yet, despite the increase in ADP concentrations, the assembly pathway is nearly identical to the "Wild-type" sample in ATP, further suggesting that ATP/ADP ratios lower than 1 are not necessary to mimic assembly in all ADP.

As mentioned above, tetramer accumulation is also observed in the ADP(AlF_x) condition with a maximum of 25% at 1.5 μM (**Fig. 2.10E**). However, unlike "Wild-type", we observe a stabilization of the hexamer over a broad concentration range, with 71% maximum occurring at 17 μM (**Fig. 2.10E**), a trend that is quite similar to assembly in the presence of ATP-γS. Additionally, the formation of larger and larger aggregates is significantly reduced, suggesting that closed-ring hexamers are favored over open-rings,

like with ATP-γS.

Assembly Mutant exhibits lower thermal stability in comparison to "Wild-type". In an attempt to better understand the effect of nucleotides on the thermal stability of both the "Wild-type" and Assembly mutant sample, we utilized previously established

Table 2.3: Thermal stability of both "Wild-type" and Assembly Mutant Rca

	"Wild-type"	*Assembly Mutant*
ADP	45.1 ± 0.5 ˚C	37.4 ± 1.1 ˚C
ATP-γS	42.7 ± 1.0 ˚C	50.9 ± 1.6 ˚C
Apo	38.2 ± 0.9 ˚C	29.7 ± 2.2 ˚C
ATP	44.2 ± 0.3 ˚C	35.9 ± 1.2 ˚C
ADP(AlF$_x$)	47.0 ± 0.4 ˚C	54.6 ± 0.8 ˚C

Experimentally determined apparent T_m values using the Thermofluor assay. Conditions testes are identical to those used in FCS, however, the protein concentration was kept constant at 5 μM. Each condition was repeated in triplicate.

Thermofluor assay methods to extract apparent melting temperatures (T_m) [22]. In brief, conditions identical to those tested in FCS were reproduced, keeping the total protein concentration fixed at 5 μM. Samples were incubated with SyproOrange, measured on a thermocycler, generating thermal melt curves (**Fig. 2.11A and 2.11B**), then analyzed to determine T_m values for all conditions tested (**Table 2.3**). Recently published work by Peterson-Forbrook *et al*. looking at subunit exchange dynamics of spinach short-form Rca showed extremely high thermal stability in the presence of ATP-γS, T_m value of 59.7 ± 0.3 ˚C, which correlated to very poor subunit exchange rates [18]. This finding lead them to concluded that this thermal stability comes from Rca being trapped in a "locked" conformation [18]. When compared to work done on "Wild-type" Rca from tobacco, there is a stark difference, as there is no significant increase in thermostability observed when in

the presence of ATP-γS (T_m value of 42.7 ± 1.0 ˚C), despite the high hexamer content observed by FCS. This would suggest that ATP-γS does not induce a "locked" conformation in tobacco Rca (**Table 2.3**). Unlike spinach Rca, the condition yielding

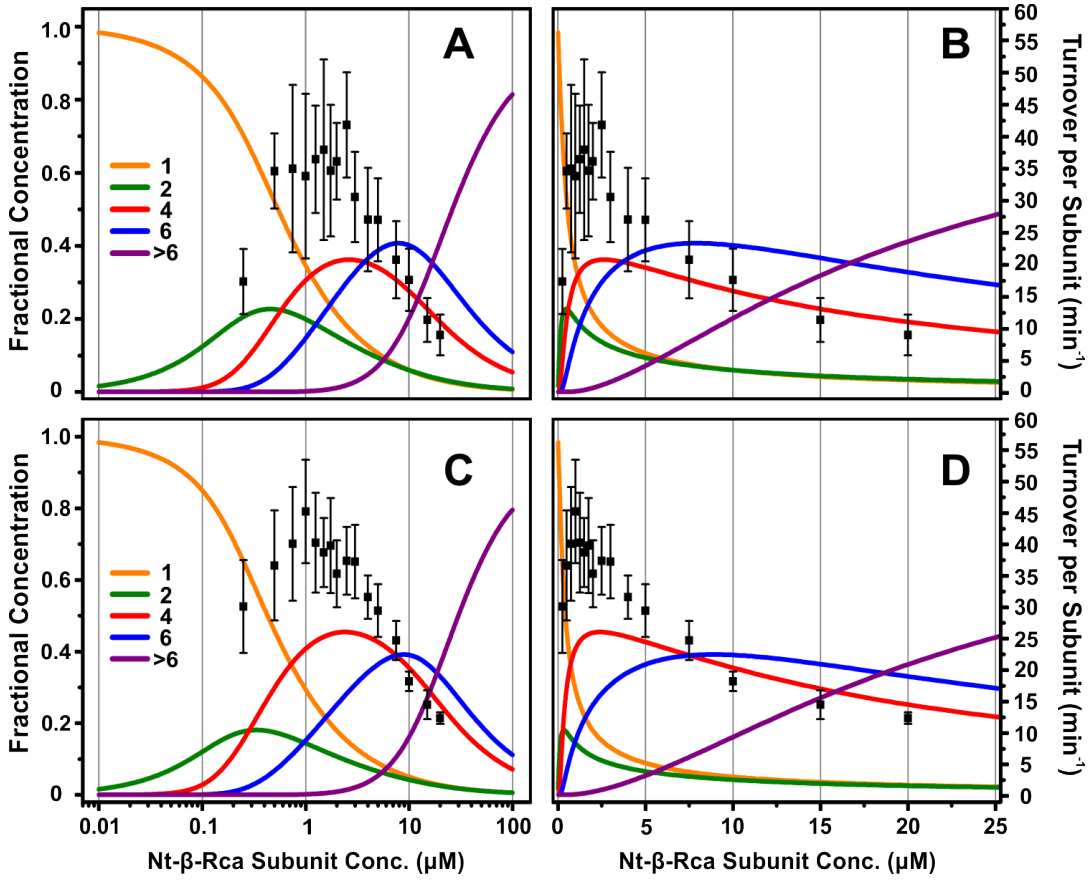


Figure 2.12: ATPase activity of both "Wild-type" and Assembly Mutant Rca. ATPase activity per subunit of both "Wild-type" Rca shown in both (A) log and (B) linear scale and the Assembly Mutant Rca shown in both (C) log and (D) linear scale. Activities are overlaid on the assembly profile from the Active ATP Turnover condition.

highest thermal stability in tobacco Rca is actually ADP(AlF_x) (47.0 ± 0.4 ˚C) (Table 3). In contrast to "Wild-type", thermal stability of the Assembly mutant does seem to be enhanced in the presence of ATP-γS (50.9 ± 1.6 ˚C), however, like "Wild-type" the highest thermal stability is actually seen in the presence of ADP(AlF_x) (54.6 ± 0.8 ˚C) (**Table 2.3**).

However, in all other conditions, there is a significant decrease in thermal stability suggesting that the Assembly Mutant is less stable than the "Wild-type" sample.

The peak ATPase activity coincides with the tetramer peak. In order to better understand the relationship between self-assembly and ATPase activity, we utilized a continuous spectrophotometric assay monitoring oxidation of NADH as a result of ADP production. Briefly, we targeted specific Rca concentrations using dilution schemes similar to those employed in FCS experiments. For each concentration, the Rca sample was mixed with assay mix and monitored at an OD_{340} until a stable signal was observed. The sample was then spiked with 2 mM of fresh ATP and monitored for an additional ~465 seconds. Production of ADP by Rca was coupled to oxidation of NADH, thus resulting in a drop in OD_{340}. The corresponding curves were then analyzed to determine the fastest initial rate over a 36 second window. These rates were then divided by the corresponding protein concentration to determine ATPase activities per subunit. It is worth noting that this assay was not limited by NADH oxidation, but rather was limited by ADP production from Rca turnover. This was determined by generating a standard curve of known ADP concentrations and ensuring that the rates measured were within the linear response region of the standard curve (**Fig. 2.5**).

Under active ATP turnover conditions, the tetramer population in the "Wild-type" sample is maximal at ~2 μM Rca, and the ATPase activity (per subunit) is maximal between 0.5 μM and 2.5 μM. Despite this match in peak position, the tetramer cannot be the sole active species because the concentration-dependent activity curve does not follow the tetramer curve (**Fig. 2.12A and 2.12B**). From peak activity to lowest activity, from 1 μM to 20 μM, there is a 76% drop in ATPase activity while only a 47% drop in tetramer

population over the same concentration range (**Fig. 2.12A and 2.12B**). This same trend is observed with the Assembly mutant, which has a maximal tetramer population at ~2 μM Rca and maximal ATPase activity (per subunit) occurring between 0.5 μM and 3 μM (**Fig. 2.12C and 2.12D**). Despite a small shift in peak activity position, like in "Wild-type", there is a 73% drop in ATPase activity from peak activity to lowest activity (1 μM to 20 μM) and only a 45% drop in tetramer population over the same concentration range (**Fig. 2.12C and 2.12D**), suggesting that activity does not follow a single oligomer. Rather, it appears that activity is highest when a mixture of dimers, tetramers, and hexamers all coexist in relatively high abundance, a condition that is met only 0.5 μM – 2.5 μM range. Therefore, dynamic interplay between all three different oligomers may be necessary for generating high catalytic rates.

Discussion:

A small amount of ADP is needed to initiate Rca's ATPase activity. In order to better understand the effect of pre-incubation of different nucleotides on ATPase activities, we prepared 3 different "Wild-type" Rca samples by buffer exchanging them into buffers containing either ADP or ATP-γS, or into buffers completely lacking any nucleotide (Apo), and then testing their activity using the previously described NADH assay. We found that incubation in different nucleotides leads to slight differences in lag times. These lag times are identified by calculating the total time required to reach the maximal rate of ATPase activity. These lag times are thought to be the time necessary for structural rearrangement of Rca in order to exchange out bound nucleotides. It is worth noting that lag times were not observed in control experiments with addition of ADP alone, suggesting that lag times are a direct result of Rca. At high protein concentrations (20 μM), incubation in ADP

yielded the shortest average lag time (51.4 ± 29.1 seconds) while Apo (73.8 ± 31.4 seconds)

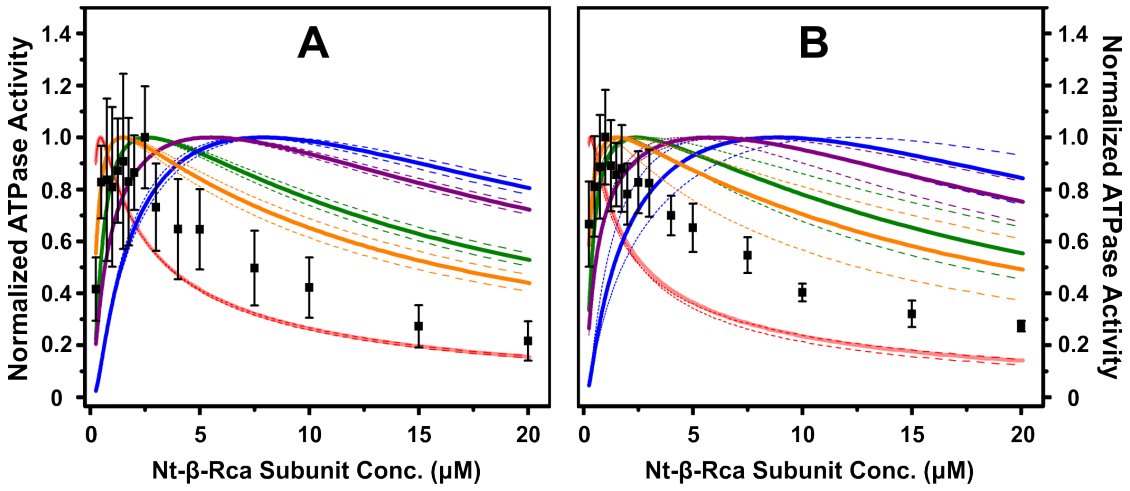


Figure 2.13: Modeling activity of the dimer, tetramer, and hexamer. This is a simulation using the fractional concentrations of the Dimer, Tetramer, and Hexamer from the Active ATP Turnover condition to test relative activities of each oligomer in order to reproduce experimentally determined activities. Both panels shows simulated data using a model in which all oligomers operate independently of each other for both (A) "Wild-type" and (B) Assembly Mutant. Different colors represent relative percent activities of each oligomer (Red, 1,0,0; Green 0,1,0; Blue 0,0,1; Purple 1,1,1; Orange 1, 0.5, 0). Dotted lines represent upper and lower limits of the modeling using the upper and lower limits of the Kd Values. All data sets have been normalized for comparative purposes.

and ATP-γS (71.2 ± 33.7 seconds) incubation resulted in slightly higher average lag times. At low protein concentrations (1 μM) the trend was less clear, given the large errors associated with each conditions (ADP 303.4 ± 107.2 seconds, Apo 311.0 ± 64.6 seconds, and ATP-γS 333.8 ± 66.6 seconds). However, despite the slight differences in lag times, one trend consistent among all three conditions was the need for a small amount of ADP to kick-start Rca's ATPase activity. For the high protein concentration (20 μM), the t=0 time point (the first second of 36 seconds used for measuring the maximal rate of ATPase activity) had an average ADP concentration of 66.4 ± 8.5 μM while the t=36 time point (the last second of 36 seconds used for measuring the maximal rate of ATPase activity)

had an average ADP concentration of 150.1 ± 31.6 μM. At low protein concentrations (1 μM), the t=0 time point had an average ADP concentration of 79.6 ± 12.8 μM while the t=36 time point had an average ADP concentration of 93.9 ± 15.7 μM. Given that the K_i value for ADP is near 40 μM [17], having just a small amount around is sufficient to initiate ATP hydrolysis. Although not being at this low of a concentration, similar ADP dependence was observed for spinach Rca when initiating subunit exchange [18].

The data support a model in which ATP hydrolysis causes dimer dissociation from the hexamer. Given that the ATPase activity did not seem to follow any particular oligomer, extensive modeling was done utilizing fractional contributions of the dimer, tetramer and hexamer from the active ATP turnover condition in order to reproduce the measured ATPase activity data. Two different models were tested. The first model assumed that each oligomer could turnover ATP independently of others which meant that reproducing the experimentally determined activity curve was a matter of summing up the relative activities of each oligomer at each concentration. The second model assumed that ATPase activity is rate-limited by the association of the dimer and tetramer into a closed ring hexamer.

Model 1 Simulation: In brief, each subunit of each oligomer over the entire concentration range (0.25 to 20 μM) was given the same relative activity of 1, using the fractional concentrations of the dimer, tetramer, and hexamer, from the Active ATP Turnover condition. Because each subunit was given the same activity and not the oligomer, the tetramer had twice the activity of the dimer, and the hexamer had three times the activity of the dimer and 1.5 times the activity of the tetramer. All activities for each oligomer were then summed up and normalized to 1. In order to confirm the notion that

one single oligomer could not reproduce the activity data, we modeled just the dimer (relative activity of dimer is 1, tetramer is 0, and hexamer is 0) shown in pink (**Fig. 2.13A**) (red dashed lines above and below show error estimates calculated from error estimates determined from different K_d values), just the tetramer (relative activity of dimer is 0, tetramer is 1, and hexamer is 0) shown in green (**Fig. 2.13A**), and just the hexamer (relative activity of dimer is 0, tetramer is 0, and hexamer is 1) shown in blue (**Fig. 2.13A**), all of which failed to reproduce the measured activity data. Given that one single oligomer could not reproduce the data, we modeled a simple summation of all three oligomers, where all subunits in each oligomers were equally active (relative activity of dimer is 1, tetramer is 1, and hexamer is 1) shown in purple (**Fig. 2.13A**). This too failed to reproduce the data. After trying different combinations of possible activities, the one yielding best results was one where the dimer and tetramer had the same activity per oligomer (tetramer had half the activity per subunit) and the hexamer had no activity (relative activity of dimer is 1, tetramer is 0.5, and hexamer is 0) shown in orange (**Fig. 2.13A**). This model, if correct, would suggest that in some way, oligomerization appears to induce negative cooperativity. This same trend was observed from identical work carried out using fractional concentrations of all three oligomers of the Assembly Mutant (**Fig. 2.13B**). Greater variability is observed with the Assembly Mutant modeling due to larger error estimates on the K_d values determined for this condition. Despite approximating the tops of the errors bars of the ATPase activity, one limitation of this model is that it eliminates all activity of the hexamer. Because other activases within the family function as hexamers, as well as other AAA+ proteins in the larger family, this model seems less physiologically feasible.

Model 2 explanation: The second model tested involved using the fractional concentrations of both the dimer and the tetramer from the Active ATP Turnover condition. At 2.5 µM, significant amounts of dimers, tetramers and hexamers coexist, and a spike in

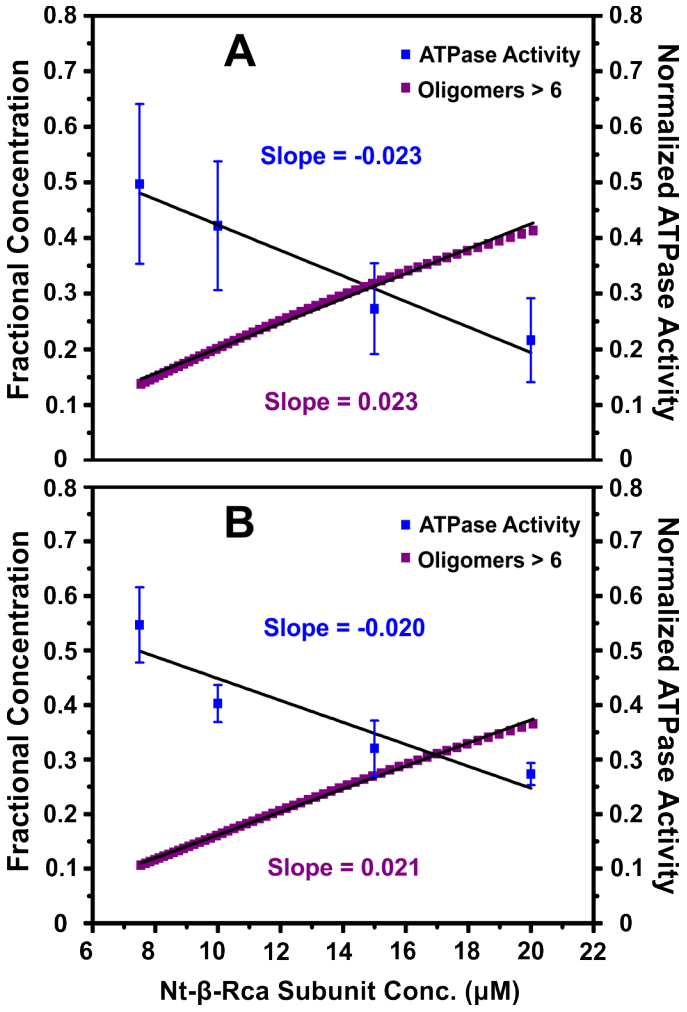


Figure 2.14: ATPase activity vs. Aggregate accumulation. Comparison between the decrease in ATPase activity and formation of oligomers >6 for both (A) "Wild-type" Rca and (B) Assembly Mutant Rca.

catalytic rates is observed (**Fig 2.12A**), however, the subsequent drop in activity at concentrations above 3 µM may be a consequence of reduced dynamic exchange between dimers, tetramers and hexamers, as the dimer population is diminished, which reduces the

propensity to assemble into hexameric rings. This model, in contrast to model 1, has more physiological relevance, given that it does not eliminate the role of the hexamer in turning over ATP, and that it allows for cycling through the different oligomers in response to nucleotide hydrolysis. Similar comparisons done with the Assembly Mutant lead to the same conclusions, with minor differences arising from fact that the peak concentration ranges differ somewhat from "Wild-type" (**Fig. 2.13B**)

Assemblies larger than hexamers are catalytically inactive. In combination, the Rca self-association and activity data suggest that species larger than hexamers are devoid of ATPase activity. This is quantitatively illustrated at concentrations between 7.5 and 20 μM, where turnover rates decrease steeply and aggregates larger than hexamers accumulate rapidly (modeled as oligomers larger than hexamers in **Fig. 2.14A and 2.14B**). Data used in this comparison comes from the active ATP turnover condition for both the "Wild-type" (**Fig. 2.14A**) and the Assembly Mutant (**Fig. 2.14B**). Upon normalization, a perfect match is observed of the rate of activity loss and the rate of aggregate formation, with respective slopes of -0.023 and +0.023 extracted from a linear fit (**Fig. 2.14A**). A similar trend is observed for the Assembly Mutant sample, where the slope of activity loss (-0.020) and the slope of aggregate formation (+0.021), are quite similar (**Fig. 2.14B**). These data provide strong support for the notion that higher-order oligomers (>Rca6) are catalytically inactive, and may serve as a storage mechanism for Rca to prevent wasteful turnover.

Purposed Rubisco reactivation model involves cycling through different oligomeric states: Based on the work presented here, the proposed rubisco reactivation model utilizes a closed-ring hexameric assembly to bind to the large subunit of Rubisco substrate. It is likely that the hexamer binds to Rubisco because the assembly in the

presence of ATP-γS promotes hexamerization and is considered to be the tight-binding pre-hydrolysis state [18]. ATP hydrolysis within a closed-ring hexamer will lead to the opening of the ring into a lock-washer state, and subsequently resulting in the dissociation of dimer from hexamer, generating the tetramer while conformationally remodeling Rubisco to

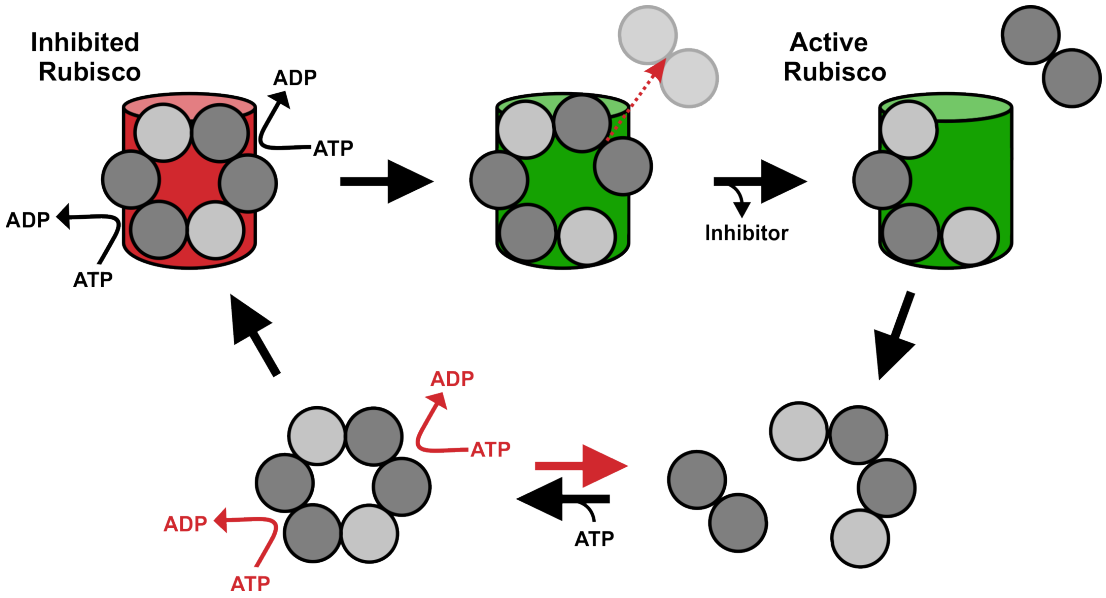


Figure 2.15: Proposed Reactivation mechanism for inhibited Rubisco. Based on the abundance of the hexamer in the ATP-γS (pre-hydrolysis) state, this work suggests that the closed-ring hexameric species binds to the inhibited Rubisco. Upon hydrolysis of ATP, the ring opens into a lock-washer conformation, and a dimer subsequently dissociates from Rubisco, providing the necessary amount of force to pull Rubisco's C-terminal tail and allow for release of the inhibitor. The tetramer then dissociates from Rubisco yielding activated Rubisco. Once dissociated, the dimer and tetramer can reassemble into a hexamer and repeat the reactivation process. In the absence of Rubisco, ATP hydrolysis by a closed-ring hexamer leads to ring-opening and disassembly into a dimer and tetramer, with subsequent assembly back to the hexamer.

release inhibitors (**Fig. 2.15**). With the inhibitor removed, tetramers and dimers may diffuse through the stroma, then re-associate to hexameric species for continued Rubisco reactivation (**Figure 2.15**). As Rubisco outnumbers Rca substantially in the chloroplast (Rubisco : Rca_6 ≈ 10 : 1), Rca_6 has a higher likelihood to bind to Rubisco rather than to self-associate. In addition, the highly crowded environment may also favor Rca_6 / Rubisco

interactions over self-assembly. Given that Rca can hydrolyze ATP in the absence of Rubisco, this same assembly/disassembly process may still occur until an inactivated Rubisco happens to be nearby, explaining the significant accumulation of all three oligomers in the FCS work. One could fathom that re-association of the dimer and tetramer would be enhanced if it were to occur on Rubisco itself, reducing the need to collide with each other in an appropriate orientation, however, we do see a significant of hexamer formation independent of Rubisco, suggesting that assembly does not take place on Rubisco. In addition, work carried out by Hazra *et al* has shown that Rca ATPase activity does not increase when in the presence of Rubisco, further supporting the notion that assembly does not occur on Rubisco [17]. This purposed model provides a good rationale for the polydisperse nature of Rca by relating it to the fundamental reactivation mechanism of Rubisco.

Conclusion:

One of the greatest challenges facing the green-type activase world today is the inability to correlate assembly of Rca with its activity. Much of the work done thus far has only been able to compare average molecular weights and average activity [23]. This comparison can be a serious shortcoming, as it may not actually report on the overall distribution of each oligomer, making it difficult to identify the oligomeric state of the active oligomer. However, work presented here shows for the first time a systematic and comprehensive characterization of the effects that different nucleotides exhibit on tobacco Rca self-assembly and a comparison of this self-assembly with ATPase activity. Additionally, we show that large aggregates are catalytically inactive and propose that they may serve as a storage mechanism of Rca in the dark, when the energy charge of the stroma

is low. However, because Rubisco outnumbers Rca, Rca hexamers have a higher likelihood of binding to Rubisco rather than to self-associate, minimizing the formation of these oligomers inside the chloroplast.

References:

1. Bracher, A., *et al.* 2015. Degradation of Potent Rubisco Inhibitor by Selective Sugar Phosphate. *Nat. Plants* **1,** 14002. DOI: 10.1038/NPLANTS.2014.2

2. Parry, M. A., *et al*. 2013. Rubisco Activity and Regulation as Target for Crop Improvements. *J. Exp. Biol*. **64,** 717-730

3. Portis, Jr. A. R., *et al.* 2008. Regulation of Rubisco activase and its Interaction with Rubisco. *J. Exp. Bot.* **59,** 1597-1604.

4. Carmo-Silva, A. E. and Salvucci, M. E. 2013. The Regulatory Properties of Rubisco Activase Differ Among Species and Affect Photosynthetic Induction During Light Transitions. *Plant Physiol.* **161,** 1645-1655

5. Bracher, A., *et al.* 2017. Biogenesis and Metabolic Maintenance of Rubisco. *Annu. Rev. Plant Biol.* **68,** 29-60.

6. Bhat, J. Y., *et al.* 2017. Rubisco activases: AAAplus chaperones adapted to enzyme repair. *Front. Mol. Biosci.* **4,** Article 20.

7. Sutter, M. *et al.* 2015. Structural Characterization of a Newly Identified Component of α-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Scientific Reports.* **5,** 16243. DOI: 10.1038/srep16243

8. Tsai, Y. C., *et al.* 2015. Identification and Characterization of Multiple Rubisco Activases in Chemoautotrophic Bacteria. *Nat. Commun.* **6,** 8883. DOI:10.1038/ncomms9883

9. Sharkey, T. D. and Zhang, R. 2010. High temperature effects on electron and proton circuits of photosynthesis. *J. Integr. Plant Biol.* **52,** 712-722.

10. Perdomo, J. A., *et al.* 2017. Rubisco and Rubisco activase play an important role in the biochemical limitations of photosynthesis in rice, wheat and maize under high temperature and water deficit. *Front. Plant Sci.* **8,** Article 490.

11. Stotz, M., *et al*. 2011. Structure of Green-type Rubisco activase from Tobacco. *Nat. Struct. Mol. Biol.* **18,** 1366-1370.

12. Chakraborty, M. *et al.* 2012. Protein Oligomerization Monitored by Fluorescence Fluctuation Spectroscopy: Self-Assembly of Rubisco activase. *Biophys. J.* **103,** 949-958.

13. Kuriata, A. M. *et al.* 2014. ATP and Magnesium Promote Cotton Short-Form Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Hexamer Formation at Low Micromolar Concentrations. *Biochemistry* **53,** 7232-7246.

14. Hasse, D., *et al.* 2015. Structure of Arabidopsis thaliana Rubisco activase. *Acta Cryst.* **D71,** 800-808. DOI: 10.1107/S1399004715001182

15. Li, C., *et al.* 2005. Two Residues of Rubisco Activase Involved in Recognition of the Rubisco Substrate. *J. Biol. Chem.* **280,** 24864-24869

16. Henderson, J. N., *et al.* 2011. Atomic Resolution X-ray Structure of the Substrate Recognition Domain of Higher Plant Rubisco Activase. *J. Biol. Chem*. **286,** 35683-35688.

17. Hazra, S., *et al.* 2015. Regulation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase: Product Inhibition, Cooperativity, and Magnesium Activation. *J. Biol. Chem.* **290,** 24222-24236

18. Peterson-Forbrook, D. S., *et al.* 2017. Nucleotide Dependence of Subunit Rearrangements in Short-Form Rubisco Activase from Spinach. *Biochemistry* **56,** 4906-4921

19. Keown, J. R. and Pearce, G. 2014. Characterization of Spinach Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Activase Isoforms Reveals Hexameric Assemblies with Increased Thermal Stability. *Biophys. J.* **464,** 413-423.

20. Hersch, G. L., *et al.* 2005. Asymmetric interactions of ATP with the AAA+ ClpX6 unfoldase: Allosteric control of a protein machine. *Cel*l **121,** 1017-1027.

21. Wang, Q., *et al.* 2018. Single-molecule diffusometry reveals the nucleotide-dependent oligomerization pathways of Nicotiana tabacum Rubisco activase. *J. Chem. Phys*. **148,** 123319. DOI: 10.1063/1.5005930

22. Henderson, J. N., *et al.* 2013. Biophysical Characterization of Higher Plant Rubisco Activase. *Biochim. Biophys. Acta*. **1834,** 87-97

23. Keown, J. R., *et al.* 2013. Small Oligomers of Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase are Required for Biological Activity. *J. Biol. Chem.* **288,** 20607-20615.

24. Xu, H., *et al*. 2004. Purification and Characterization of the AAA+ Domain of Sinorhizobium meliloti DctD, a σ54-Dependent Transcriptional Activator. *J. Bacteriol.* **186,** 3499–3507

CHAPTER 3

UNDERSTANDING THE SELF-ASSEMBLY PATHWAY OF SPINACH SHORT-FORM RUBISCO ACTIVASE

Abstract:

In addition to being one of the most abundant proteins on Earth, Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) is responsible for fixing atmospheric CO_2 in plants and generating the necessary precursors for sugar production, a function critical in biomass accumulation. Although Rubisco plays such a pivotal role in plants, it is not without its setbacks. One such limitation is Rubisco's propensity to bind O_2 in place of CO_2, as well as being prone to inhibition via a variety of substrate analogs, including its own, under certain conditions. An activator protein discovered in the late 80's, termed Rubisco Activase (Rca), has been shown to reactivate Rubisco via ATP hydrolysis. In brief, Rca is a chemo-mechanical motor protein belonging to the AAA+ protein superfamily, which is thought to remove the inhibitor by modifying the conformational state of Rubisco via ATP hydrolysis. Despite its importance, the reactivation mechanism used by Rca to reactivate Rubisco is loosely understood, in part because the self-assembly mechanism for Rca is not well defined. In an attempt to address this, work present here using Fluorescence Correlation Spectroscopy (FCS) shows that for spinach short-form Rca, large aggregates are present at all concentrations. Varying incubation times and nucleotides did not seem to eliminate aggregates, however, the relative dissociation rates for these large aggregates did change with changing nucleotides. Additionally, these aggregates appeared to be dynamic in nature, as filtration through a 0.1 μm filter did not remove aggregates, indicating the ability of these large oligomers to dissociate, pass through the filter, and re-assemble on

the other side. Given the effect of varying nucleotides, it may be necessary to incubate with a mixture of nucleotides rather than a single one. Also, based on previous work form the Pearce group, incorporation of the long-form Rca may be required for proper assembly and aggregate minimization. While this work only looks at the assembly of spinach Rca, the ultimate goal is to understand and characterize the self-assembly of Rca from different species given that this assembly may vary across differing species.

Introduction:

Considered by some scholars to be the most abundant protein in the world, Rubisco [Ribuilose-1,5-bisphosphate (RuBP) Carboylase/Oxygenase] is responsible for facilitating the fixation of inorganic CO_2 from the atmosphere into precursor molecules used for sugar production in various different organisms [1,2]. It does so by combining RuBP, a 5 carbon sugar, with the incoming CO_2 molecule, and forms two 3 carbon molecules, called 3-phosphoglycerate (3PGA). This product is later processed into glyceraldehyde-3-phosphate (G3P), a molecule used by plants to generate glucose [3,4,5]. Because RuBP needs to be regenerated for the reaction to start again, the cycle needs to be completed 3 times in order to get 1 extra G3P that can be used for glucose production. This carbon fixation step of the reaction carried out by Rubisco is considered to be the bottleneck reaction of the Calvin-Benson-Bassham (CBB) cycle and is responsible for all of the biomass accumulation observed in the world today [6,7]. Because of the implications Rubisco has the bioenergy and food production sectors, a great deal of attention has been placed on better characterizing the mechanism it uses for carbon capturing, with the ultimate objective being the re-engineering of Rubisco for improved crop yields [1,8,9].

Within the Rubisco family, there are currently four different forms identified, all of which differ in both organisms they are found in and in the structural features they possess. The most common among them is Form I Rubisco, which is found higher plants and cyanobacteria as well as organisms from the red plastid lineage. Structurally, form I Rubisco is a hexadecamer, consisting of eight large subunits and eight small subunits commonly referred to as L8S8, where L stands for large and S for small [4,10]. The binding site for RuBP is generated at the interface of two anti-parallel large subunits, resulting in

two binding sites per 2 large subunits. For proper substrate turnover to occur, the active site must covalently bind a CO_2 molecule using a specific Lysine residue and have a Mg^{2+} ion in close proximity prior to having RuBP and the incoming CO_2 bind. This step is commonly referred to as carbamylation [11,12]. In addition to RuBP, there are other known substrates which can bind to Rubisco's binding site, such a CA1P, which act as inhibitors and inactivate Rubisco. Scholars believe that these inhibitors function to regulate Rubisco activity such that it is not needlessly producing 3PGA in times of low light [14]. As with CA1P and other inhibitors, binding of RuBP to the binding site prior to it being carbamylated results in the inactivation of Rubisco [15,16]. Once Rubisco is rendered inactive, a chaperone protein, termed Rubisco Activase (Rca), is recruited to restore Rubisco's activity through a process known as reactivation [13].

The Rubisco activase family consists of three distinct classes, distinguished by the lineage in which they are found [17,18]. Red-type activases have been found to reactivate form I Rubiscos from proteobacteria and members of the red plastid lineage [19,20,21]. Like some red-type activases, CbbQO-type activases are also found in proteobacteria, however, they have been shown to be capable of reactivating both Form I and Form II Rubisco [17,22,23]. The last type, and certainly the most abundant, is the green-type activases which is responsible to reactivating Form I Rubiscos from higher plants and cyanobacteria in the green plastid lineage [17,24]. Despite being prevalent in different lineages, all three types belong to the AAA+ (ATPases Associated with various cellular Activities) protein super family. Structurally, they contains the canonical AAA+ module, which houses the nucleotide binding site, along with two extensions on the N and C termini, which vary from type to type. This nucleotide binding site is located in between the two

subdomains of one single subunit and is further shielded by the interface of the adjacent subunit, which provides additional interactions that help stabilize the bound nucleotide [25,26]. Due to the interactions from the adjacent subunits, hydrolysis in one subunit is believed to be sensed by surrounding protomers, allowing large, coordinated movements to be carried out across the entire assembly [27,28]. Like most protein within the AAA+ family, Rca uses this hydrolysis of ATP to perform chemo-mechanical work on its respective Rubisco substrate to remove any bound inhibitors. Despite knowing that activases use hydrolysis in this way, a complete understanding of the mechanism used is not well understood for all activases.

Just as activases differ in the lineages they are found in, so also, the reactivation method used differs. Red-type activases, which are the best characterized activases thus far, have been shown to form obligate closed-ring hexamers in the presence of ATP and RuBP [19,20]. These closed-rings attach to the side of Rubisco and begin threading its C-terminal tail through the central pore, providing enough force to open up the active site and allow the bound inhibitor to escape. The tail is then released from the pore and the process can begin again [21]. In contrast, the CbbQO-type activases has 6 subunits assembling into a closed-ring hexamer, referred to as Q_6, and an O domain, containing the von Willebrand Factor A (VWA) domain, which is known to mitigate protein-protein interactions [17]. The O domain is tethered to the Q_6 ring in a hinge-like fashion at a stoichiometric ratio of Q_6O_1. Unlike red-type, this activase class does not thread Rubisco's tail through its pore, but rather uses the VWA domain to attach to the large subunit and pull the binding site open, allowing the substrate to be released [17,22,23].

The reactivation mechanism for the green-type activases is less well understood, in

part because mutational analysis of higher plant Rubisco is more difficult, but primarily because unlike most AAA+ proteins and other activase, green-type activases do not form obligate hexamers. Rather, they exhibits a more polydisperse character, meaning that at a given concentration, there is a distribution of oligomers rather than a single one [29,30]. What is known about the mechanism is that green-type activases can not thread the C-terminal tail of Rubisco through their central pore like red-types, because the corresponding green-type Rubisco has a tail that is shorter by about 12 residues [17]. However, in order to properly characterize this mechanism, it is first necessary to characterize the self-assembly pathway of green-type Rca.

Progress on understanding the self-assembly pathway of Rca has be gradual, in part because traditional biochemical techniques have been difficult to interpret. Great efforts were by many groups on developing and optimizing the methods for Analytical Ultra Centrifugation (AUC), and similar type experiments, in hopes of determining the predominant oligomeric state [31,32]. However, the most significant limitation of this effort was the inability of these techniques to distinguish between active and inactive species. The Pearce group used AUC experiments to compare average molecular weights, measured as a function of concentration, with ATPase activity also collected as a function of concentration. Their findings suggest that the most active oligomer, both in ATPase and reactivation activity, is something close to a trimer given that the peak activity aligned well with an average molecular weight (MW) corresponding to something larger than a trimer [31,32]. This finding is in partial agreement with work carried out by the Hayer-Hartl group at the Max Planck Institute, which also found the most active oligomer in ATPase activity to be something between a dimer and tetramer. However, in contrast, they found that the

oligomer most active in Rubisco reactivation was something between a tetramer and hexamer, making the interpretation less clear [24]. In addition to AUC, other groups have tried using size-exclusion chromatography (SEC) to characterize the distribution of oligomers present at various concentrations but have had little success due primarily to broad peaks which correspond to a complex distribution of oligomers that cannot be de-convoluted [33].

For these reasons, we have shifted towards using a fluorescence-based technique, Fluorescence Correlation Spectroscopy (FCS), which offer greater sensitivity in distinguishing different oligomeric states based on diffusion. Previous work with Rca from cotton allowed a thorough analysis of the fractional concentration of each oligomer at a given concentration. This analysis showed unique differences in assembly, depending on the nucleotide used [29,30]. Incubation with ADP resulted in the continuous assembly towards larger oligomer while incubation with ATP-γS, a non-hydrolyzable ATP analog, showed the stabilization of nearly 80% hexamers and a decrease in the propensity to assemble into larger oligomers [29,30]. This same methodology was used in this study to analyze the self-assembly pathway of Spinach short-form Rca.

The motivation behind using the short-form Rca for this study was twofold; the first reason was to compare self-assembly with work recently published on characterization of the subunit exchange dynamics as a function of nucleotide [34]. The second, was that spinach short-form Rca appeared to exhibit an unusually high thermal stability [35], especially when in the presence of ATP-γS [34]. Despite the promise that this previous work with cotton has shown, application of this FCS methodology to spinach has been only partially successful. In contrast to cotton Rca, spinach Rca appeared to have large

aggregates present at all concentrations. The severity of these aggregates was observed to vary with changing nucleotide, however, varying incubation times and temperatures did not seem to eliminate them completely. In addition, these aggregates were shown to be dynamic in nature, given that filtration immediately before experimentation did not eliminate them. This would suggest that these aggregates were able to disassemble, pass through the 0.1 μm filter, and reassemble on the other side. These finding, although not ideal, do shed more light on the complexities associated with understanding the self-assembly mechanism of Rca.

Materials and Methods:

Plasmid generation and mutagenesis. The gene encoding for So-β-Rca, residues 1-375, was removed from the pET23a(+) vector and inserted into the pHUE expression system, using the SacII and NotI restriction sites. This expression system was a gift from S. Whitney (Australian National University, Canberra, Australia) and allowed for fusion of a 6His-tagged ubiquitin tag to the spinach Rca construct, as was done with Tobacco Rca. Doing this allowed the fusion construct to be cleaved by a 6His-tagged deubiquitylating enzyme (DUB) after purification. Like with tobacco, the strategy for specifically labeling the protein with a fluorescent dye involved the replacing of an existing threonine with a cysteine residue at position 375 (T375C) using the QuikChange site-directed mutagenesis kit (Agilent) and provided instructions. The cysteine substitution was verified by DNA sequencing.

Transformation and Growth. The same methodology used for the tobacco Rca construct described previously was used for the spinach Rca construct. In brief, the pHUE vector containing Rca gene was transformed into *Escherichia coli* strain BL21*(DE3)

(Invitrogen) using the heat-shock method. Single colonies were picked and grown overnight (14 – 16 hours) in LB media containing 100 μg/mL carbenicillin. Like tobacco, overnight cultures were used to inoculate a (6) 1L cultures and grown at 37 ˚C and 250 rpm, until the OD_{600} reached 0.6. They were then induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown for an additional 8 – 10 hours at 25 ˚C and 200 rpm, after which they were harvested and frozen at -80 ˚C.

Protein Purification. As was previously described for tobacco Rca, the cell pellet containing spinach Rca (further referred to as So-β-Rca) was suspended in 100 mL of Solubalization buffer and stirred for 15 minutes at 4 ˚C. It is worth noting that initial purification conditions for spinach Rca involved use of 0.1 mM ADP instead of 0.5 mM. However, the protocol was changed to include the higher nucleotide content, like tobacco, because purified Rca was less aggregate prone / misfolded than with less nucleotide. Resuspended cells were then lysed by sonication and centrifuged for 30 minutes. The supernant was then passed through a syringe filter and loaded onto a HisTrap HP nickel-nitrilotriacetic acid (Ni-NTA) column (GE) on the ÄKTA pure FPLC system (GE) for purification by affinity chromatogaphy, like tobacco Rca. In brief, the HisTrap HP column was washed with 10 mM Imidazole to eliminate unbound species. A step gradient to 100 mM Imidazole was then used to wash off weakly bound species. Like tobacco, Spinach Rca was eluted with a linear gradient from 100 mM to 500 mM Imidazole (Rca eluted at 200 mM Imidazole). After pooling all fractions containing So-β-Rca, deubiquitylating enzyme (DUB) was added to the sample and it was dialyzed overnight at 4 ˚C against 1 L of Dialysis buffer. After dialysis, the sample was run over a HisTrap HP column for a second time and, due to the cleavage of the 6His-tagged Ubiquitin tag, eluted in the early

fractions. The sample was then pooled and applied to a HiTrap Q HP 5mL anion exchange column (GE) and washed with 40 mM NaCl to remove unbound material. A step gradient to 160 mM NaCl was then used to wash off weakly bound species. Desired So-β-Rca was eluted with a linear gradient from 160 mM to 500 mM NaCl (Rca eluted at 250 mM NaCl). All fractions containing So-β-Rca were pooled and concentrated down to 5 mL using a Centriprep concentrator (Millipore, Milford, MA). Unlike tobacco Rca, samples were then buffer exchanged using a method termed "Soft buffer exchange" in place of the PD-10 columns (GE Healthcare). This "Soft buffer exchange" method involved mixing a desired buffer with the concentrated sample, to a total volume of 15 mL (10 mL of target buffer and 5 mL of concentrated protein) inside the concentrator. The new mixture was then concentrated back down to 5 mL volume, at which time, the cycle started again. This cycle was repeated until about 40 mL of target buffer had been used. It is worth noting that many attempts were made to try other concentrating methods, including the PD 10 column and Ammonium Sulfate precipitation, however, both lead to aggregated protein that proved problematic when used in subsequent assays. After the 40 mL had been used, the sample was concentrated down to 1 mL total volume, aliquoted, flash frozen in liquid nitrogen, and stored at -80 ˚C. Protein concentrations were determined using the Bradford method (n = 9) using BSA as a standard, with typical yields of 1.5-2.5 mg/L of cell culture. Fractions from each of the various purification steps, as well as the final sample, were analyzed by SDS-PAGE to determine purity, which ranged from 85 – 98% (determined visually).

Labeling So-β-Rca. The labeling strategy for So-β-Rca was identical to that of tobacco and involved targeting the engineered cysteine residue at position 375 of the C-

terminal tail with Alexafluor dyes functionalized with a maleimide group on the 6-carbon linker. In brief, generating Alexa Fluor 546 C5-maleimide ($\lambda_{max\text{-}ex}$ = 554 nm; $\lambda_{max\text{-}em}$ = 570 nm; MW = 1034 Da) (Molecular Probes, Inc.) labeled So-β-Rca involves mixing both dye and Rca at a molar ratio of 2 to 1 (dye to protein) in the presence of ATP-γS. It is worth noting that this ratio was reduced significantly from 3.5 to 1, which resulted in extreme over-labeling. Like tobacco, the total protein concentration was kept at 100 μM along with a dye concentration of 200 μM, in a total volume of 300 μL. The reaction was then allowed to proceed for 4 hours at 4 ˚C, which is shorter than the previously established protocol and helped reduce the amount of over-labeling. To remove excess free dye, 300 μL of saturated ammonium sulfate was added incrementally to the sample and incubated at 4 ˚C for an additional 30 minutes. The sample was pelleted down and desalted over 2 gel filtration spin columns pre-equilibrated with the buffer. Desalted sample was then aliquoted, flash frozen in liquid nitrogen, and stored at -80 ˚C.

Rca Characterization by High-Performance Liquid Chromatography (HPLC) and Matrix Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometry. As with tobacco, both label and label-free samples were analyzed by reverse-phase HPLC (Agilent Technologies 1100 Infinity Quaternary LC system, Agilent 1260 Series Diode-Array detector) using a linear water/acetonitrile gradient with 0.1% trifluoroacetic acid. The absorbance of the sample was monitored at OD_{220}, OD_{280}, and OD_{554} for Alexa Fluor 546, to ensure that both protein and dye would be detected. The chromatogram for the unlabeled spinach Rca contained a single peak (elution at 47.3% B for label-free sample), however, unlike tobacco, the chromatogram of the labeled sample contained a main peak with a smaller secondary peak (elution of main peak at 48.3% B for Alexa Fluor 546

labeled sample, elution of secondary peak at 49.1% B for Alexa Fluor 546 labeled sample), suggesting a certain degree of over-labeling. Despite the over-labeling, the chromatogram did lack any free dye peaks suggesting that all dye present was covalently attached to the protein.

As with tobacco, labeling efficiencies were determined from the absorbance of each species, using the same methodology described previously. In brief, the "Wild-type" Rca concentration was determined using the absorbance at OD_{280} and the calculated extinction coefficient (ε_{280} = 39,880 M^{-1} cm^{-1}). The OD_{554} correction factor for Alexa Fluor 546 dye was 0.12 and was subtracted from the OD_{280} signal.

In addition to HPLC, labeled and label-free samples were also analyzed by MALDI mass spectrometry (Applied Biosystems/Voyager DE STR MALDI TOF/TOF analyzer) using sinapinic acid as the matrix. The experimentally determined mass for label-free spinach Rca was 41,891 Da, while that carrying the Alexa Fluor 546 was 42,894 Da (theoretical mass values for label-free and labeled samples are 41,926 Da and 42,960 Da respectively with the theoretical mass of Alexa Fluor 546 being 1,034 Da). However, it is worth noting that despite having a peak corresponding to a singly labeled sample, the MALDI-TOF spectrum did contain a poorly defined shoulder peak that could be attributed to a doubly labeled sample.

FCS Experiments and Data Analysis. Prior to beginning work on the home-built Fluorescence Correlation Spectroscopy (FCS) instrument, glass slides were thoroughly cleaned using the previously described method. In brief, glass slides were treated with Ozone gas, sonicated in 3% Hellmanex (Hellma) cleaning solution, then washed with water and dried with nitrogen gas. After slides were cleaned, FCS work began by powering on

the laser and allowing a 20 minute warm up period, during which time the glass coverslip and perfusion chamber were prepared and placed on the setup, as previously described. It is worth noting that initial BSA treatment of the chamber involved use of a 1X solution (0.1 mg/mL). However, with this amount of BSA, the spinach Rca sample was seen to have greater interaction with the glass slide, thus resulting in a need for more BSA. This finding suggests that Rca from spinach, unlike tobacco and cotton, experiences greater degree of structural rearrangement, which results in the exposure of more hydrophobic regions. Therefore, the new protocol involved treatment of the perfusion chamber with 5X BSA (0.5 mg/mL). Once treated, a 20 nM sample of TAMRA ($\lambda_{max\text{-}ex}$ = 548 nm; $\lambda_{max\text{-}em}$ = 574 nm; MW = 430 Da) (Molecular Probes, Inc.), a free dye solution with a known diffusion coefficient (420 $\mu m^2 s^{-1}$), was used to determine confocal parameters, as described previously for tobacco work.

Once instrumental preparations were complete, the label and label-free samples of Nt-β-Rca were removed from the -80 ˚C freezer and thawed on ice for 10 minutes. An identical dilution scheme was used for the spinach work as was previously described for tobacco Rca. Briefly, a series of stock solutions were prepared for both labeled and label-free samples using buffer containing the particular nucleotide of interest and incubated at varying temperatures, depending on the aims of the experiments. Temperatures ranged between 4˚C and 23˚C. The concentration range tested for the spinach Rcs work ranged from 0.05 μM to 45 μM, with particular emphasis in the low μM region. After generating the appropriate concentration, 35 μL of the sample was placed inside the perfusion chamber and measured. This procedure was then repeated for all concentrations in all conditions. Like tobacco, a purposeful attempt was made to prevent systematic errors in the data

collection process, thus protein concentrations were collected in various orders, ensuring that each concentration experienced the same average incubation time in each particular condition. The raw autocorrelation curves for each concentration were analyzed individually to identify and remove any and all traces containing aggregated material. The remaining traces were then averaged and fit with the autocorrelation function to extract apparent diffusion coefficients. The confocal parameters used for fitting were those extracted from fitting the free dye at the beginning of the experiment. Because data collection was conducted over several months, r values were determined independently for each experiment. All fits were done manually and inspected visually.

Results:

Large aggregates are present at all concentrations. In an attempt to better characterize the assembly of spinach short-form Rca as a function of nucleotide, we have used previously established FCS methods to monitor its' assembly as a function of concentration [29,30]. As with tobacco, protein preparation for FCS experiments involved purification of a recombinantly expressed fusion protein in *E. coli* containing short-form Rca and an N-terminal 6His-Ubiquitin tag, by Ni-affinity and Anion-exchange chromatography. Removal of the tag was carried out by incubation with a deubiquitylating enzyme (DUB). As with Cotton [29,30] and tobacco, the strategy for attachment of a fluorescent labeled involved use of an engineered Cys residue on the C-Terminal tail, incorporated via an Thr-Cys substitution at the 375 position (spinach numbering), thus providing a site for thiol directed labeling of the AlexaFluor-546 by means of maleimide chemistry. Analysis by HPLC and MADLI-TOF confirmed overall successful incorporation of a single dye per subunit of Rca, with molar dye to protein ratios of 1.4 ±

0.2 : 1 (n=4). These labeling efficiencies are higher than those observed for the tobacco constructs, suggesting a greater degree of over-labeling. As with tobacco Rca, apparent (average) diffusion coefficients (D_{app}) were measured on a small number of particles, with

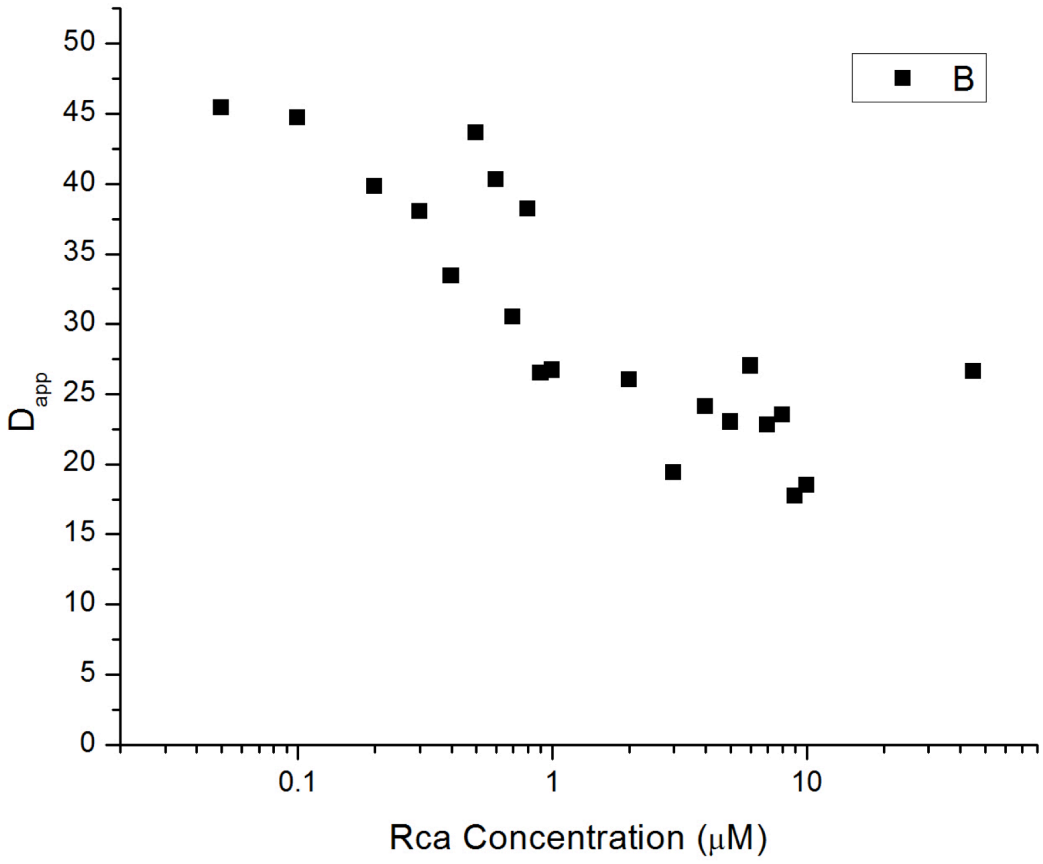


Figure 3.1: Extracted diffusion coefficients for Spinach short-form Rca. D_{app} values extracted from fits of the FCS data for each of the concentration measured. They are not normalized with respect to any oligomer.

subunit concentrations of label-free Rca ranging from 0.05 to 45 μM and dye-labeled Rca concentrations being kept constant at 0.05 μM.

Unlike previous work with tobacco, FCS measurements showed a high abundance of aggregated protein carrying multiple labels per assembly and thus distorting the resulting autocorrelation curve. As such, these aggregates resulted in the removal of, in most cases,

50% of the collected data. The abundance of these large oligomers did not seem to change with increasing concentrations, suggesting some sort of equilibrium between small and

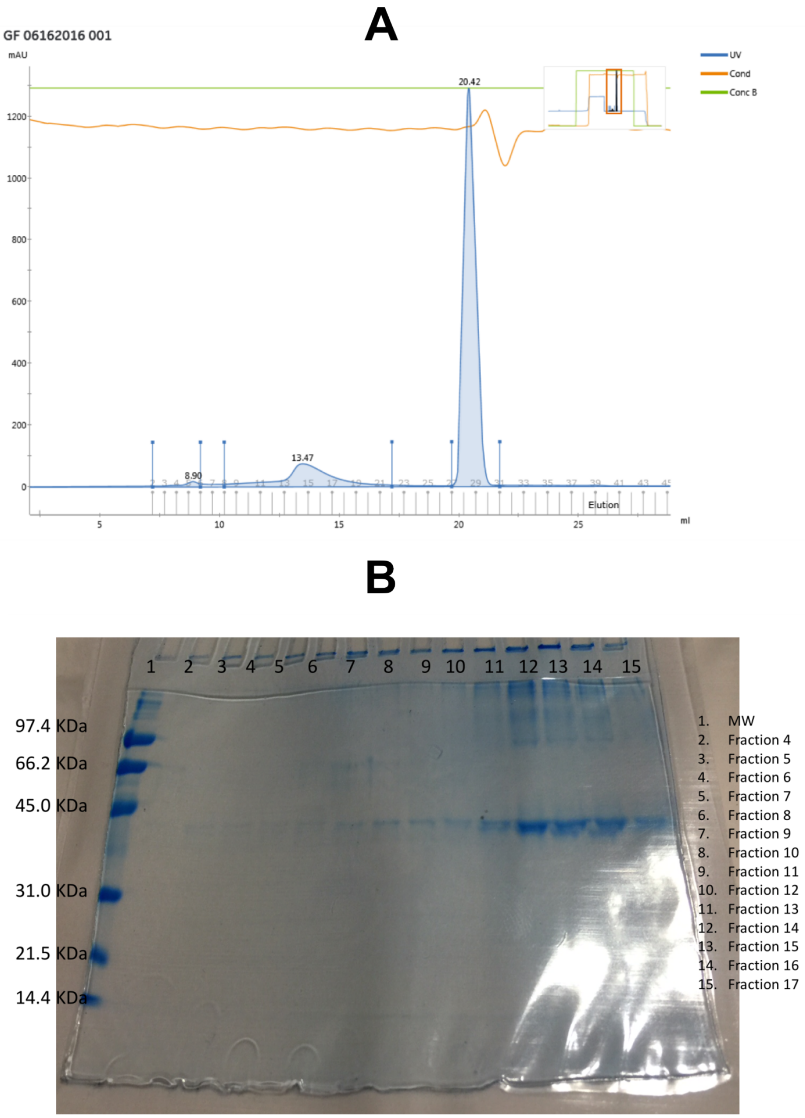


Figure 3.2: Size Exclusion Chromatography (SEC) on Spinach short-form Rca. (A) Size exclusion chromatogram of Spinach short-form Rca showing broad peak over Fractions 12 through 21. The sharp peak at fraction 29 was not protein, as confirmed by Bradford. Each fraction was analyzed by a (B) SDS-PAGE gel to determine the MW of the sample. Each fraction contained Rca.

large oligomers. Using the poor quality FCS data, an assembly curve was generated (**Fig. 3.1**), showing a decrease in the apparent diffusion coefficient (D_{app}) indicating an overall increase in the average size of the diffusing oligomer with increasing concentration. It is

worth noting that due to the poor quality of the FCS data, the resulting fit using the autocorrelation function is also poor. Therefore, much weight cannot be placed on the accuracy of the measured D_{app} values.

Along with FCS, additional experiments were carried out using size exclusion

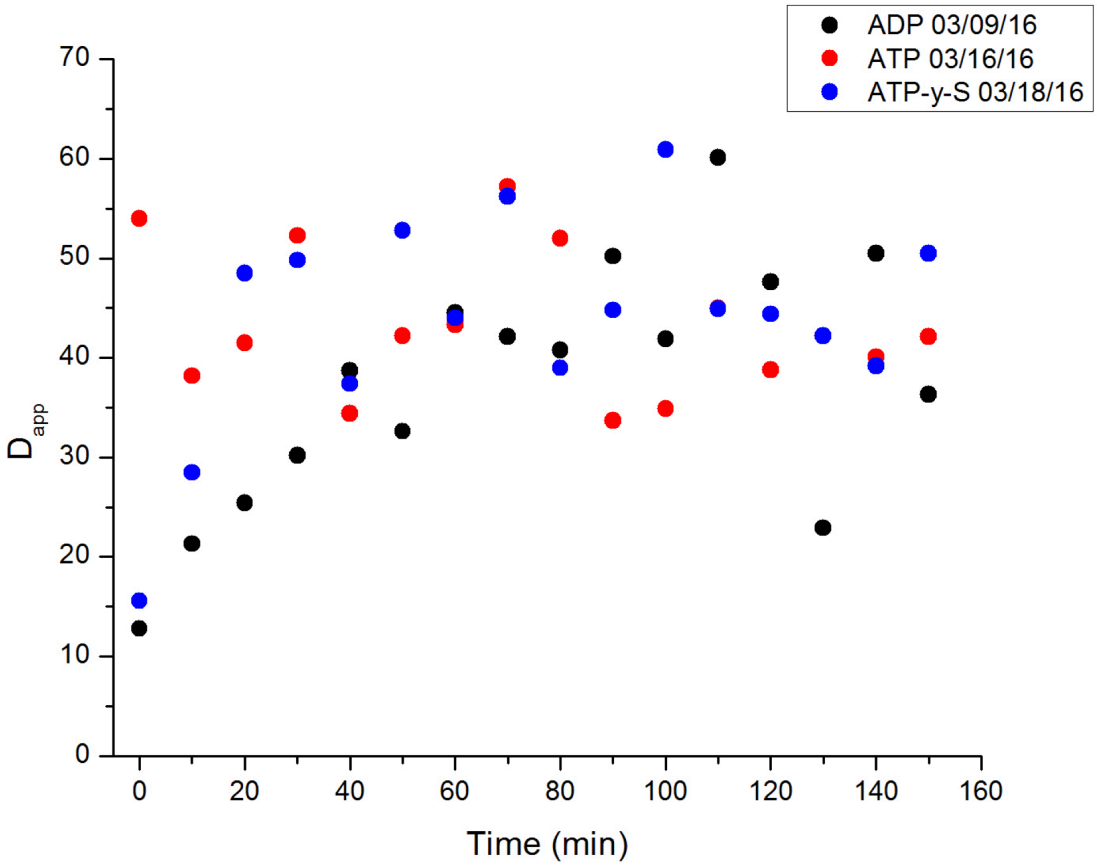


Figure 3.3: Nucleotide effect on disassembly over time. D_{app} values for 0.05 μM as a function of time. These measurements were collected on three different days in the presence of different nucleotides, either ADP (black data), ATP (red data), or ATP-γS (blue data).

chromatography (SEC) and dynamic light scattering (DLS) to address the aggregation issue. Work done on SEC shows a broad peak on the chromatogram with an extended tail rather than a typical sharp and distinct peak (**Fig. 3.2A**). This broad trailing peak was also

observed previously for the cotton Rca construct. When analyzed by SDS-PAGE, it was found that all fractions within the broad peak contained Rca, suggesting that different sizes of Rca were eluting at different elution times (**Fig. 3.2B**). Additionally, these same fractions were also analyzed by DLS where it was determined that like FCS, all fractions contained a mixture of really small oligomers and really larger aggregates.

Different nucleotides show distinct differences in assembly. Given the severe aggregation issue, we decided to monitor assembly by FCS using different nucleotides. However, instead of trying to monitor assembly as a function of concentration, as with tobacco, we first wanted to examine whether different incubation times were necessary to allow for disassembly of these large aggregates. It is worth noting that initial experiments carried out kept all stock solutions at 4˚C, however, doing so, resulted in extremely poor data, prompting a change in incubation temperature for all stock solutions to ambient room temperature (RT), about 23 ˚C.

At a fixed concentration of 0.05 μM, incubation with ATP appeared to give mostly monomeric assemblies at short incubation times (**Fig. 3.3**). Despite not having a distinct apparent diffusion coefficient for the spinach Rca monomer, observed D_{app} values are consistent with those determined for both tobacco and cotton. Over the course of 2.5 hours (measurements taken every 10 minutes), the So-β-Rca, remained primarily monomeric with slight fluctuations in measured D_{app} values, suggesting minor changes in assembly state. In contrast, incubation with ADP initially gave very large assemblies at short incubation times (**Fig. 3.3**). If an average monomer D_{app} value were to be used for comparative purposes, a value of 55 $\mu m^2 s^{-1}$ which is in line with those identified for tobacco and cotton constructs, the resulting size of the oligomer at short incubation times

in ADP would be a something close to a 72-mer. However, over the course of 1 hour,

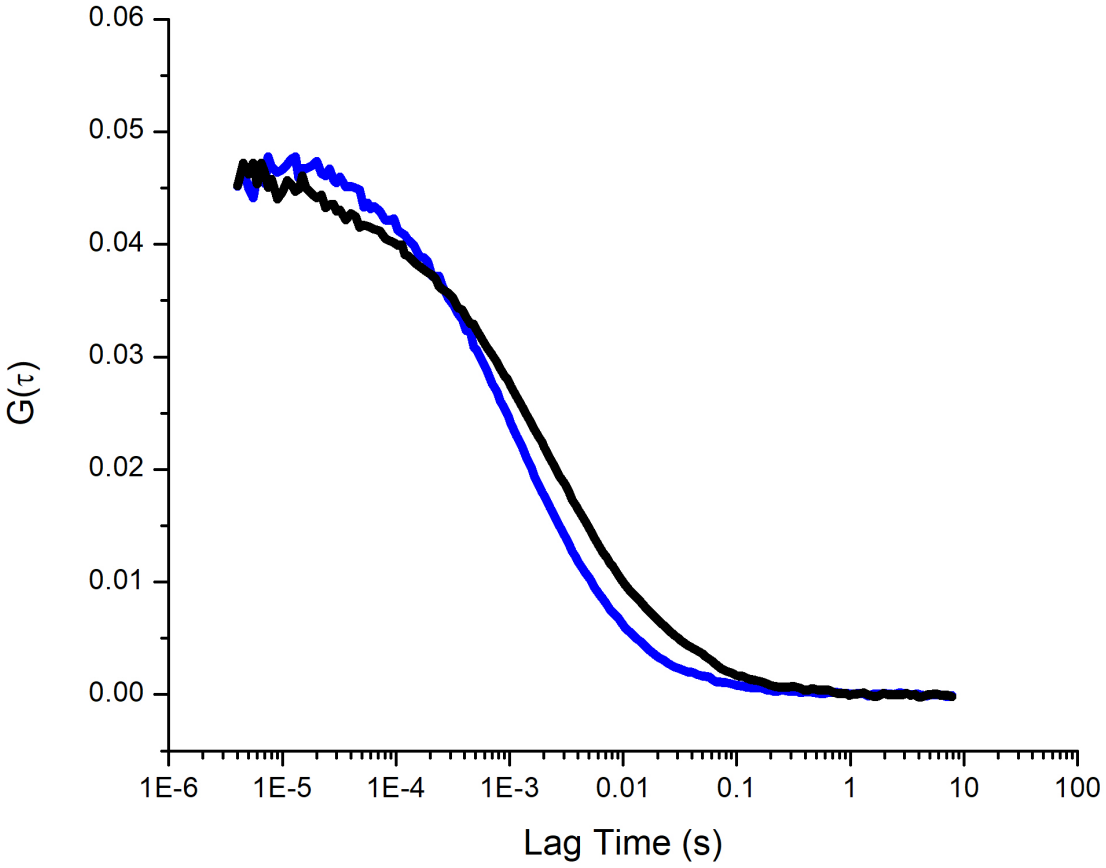


Figure 3.4: Autocorrelation curve shape from FCS work. Comparison between the autocorrelation curve of spinach Rca in the presence of ADP (blue) and ATP-γS (black), normalized with respect to one another.

drastic disassembly is observed into a D_{app} value consistent with a dimer. At times longer than 1 hour, the measured D_{app} values generally remained constant (**Fig. 3.3**). Like ADP, examination with ATP-γS, a non-hydrolysable ATP analog, showed similarly sized large oligomers at short incubation times (**Fig. 3.3**). Within 20 minutes, So-β-Rca appeared to disassemble into much smaller oligomers, and remained nearly constant over the remaining 2 hours (**Fig. 3.3**). However, it is worth noting that the autocorrelation decays and

corresponding fits for this condition were very poor due to a coexistence of both small and large oligomers. Typical decay curves have a flatter region at short lag times but soon become more rounded in their decay to zero as seen with the ADP condition (**Fig. 3.4 blue line**). In contrast, the decay curve of So β Rca incubated with ATP-γS shows an almost linear decay over the course of the measurement (**Fig. 3.4 black line**), a trend that is typically associated with distributions of both very small and very large particles. Therefore, despite yielding D_{app} values consistent with smaller oligomers, assembly in the presence of ATP-γS severely halts the disassembly of large oligomers. This trend is consistent with recently published data on So-β-Rca which shows poor subunit exchange dynamics when in the presence of ATP-γS [34]. The authors speculate that incubation with ATP-γS induces a tightly-bound state essentially locking the oligomer in whatever assembly it is. Given the sample measured was generated from a highly concentrated stock solution, it is not surprising to find larger oligomers still present in solution. Because of this distribution issue, further experiments were not carried out with ATP-γS,

In order to identify whether this nucleotide effect is maintained at higher concentrations, we performed a similar experiment as before at a concentration of 4 μM instead of 0.05 μM. As observed prior, incubation with ATP yields D_{app} values that are nearly constant at all incubation times (**Fig. 3.5**). In contrast, incubation with ADP appears to exhibit the same disassembly trend seen at the lower concentration, suggesting that in the presence of ADP only, there is a delay in the disassembly process (**Fig. 3.5**).

Large aggregates are dynamic in nature. Given that differing nucleotides and incubation times did not seem to solve the aggregation problem, we speculated that those remaining aggregates may be a byproduct of the purification process and must thus be

misfolded protein that has associated together into one large oligomer. To test this theory,

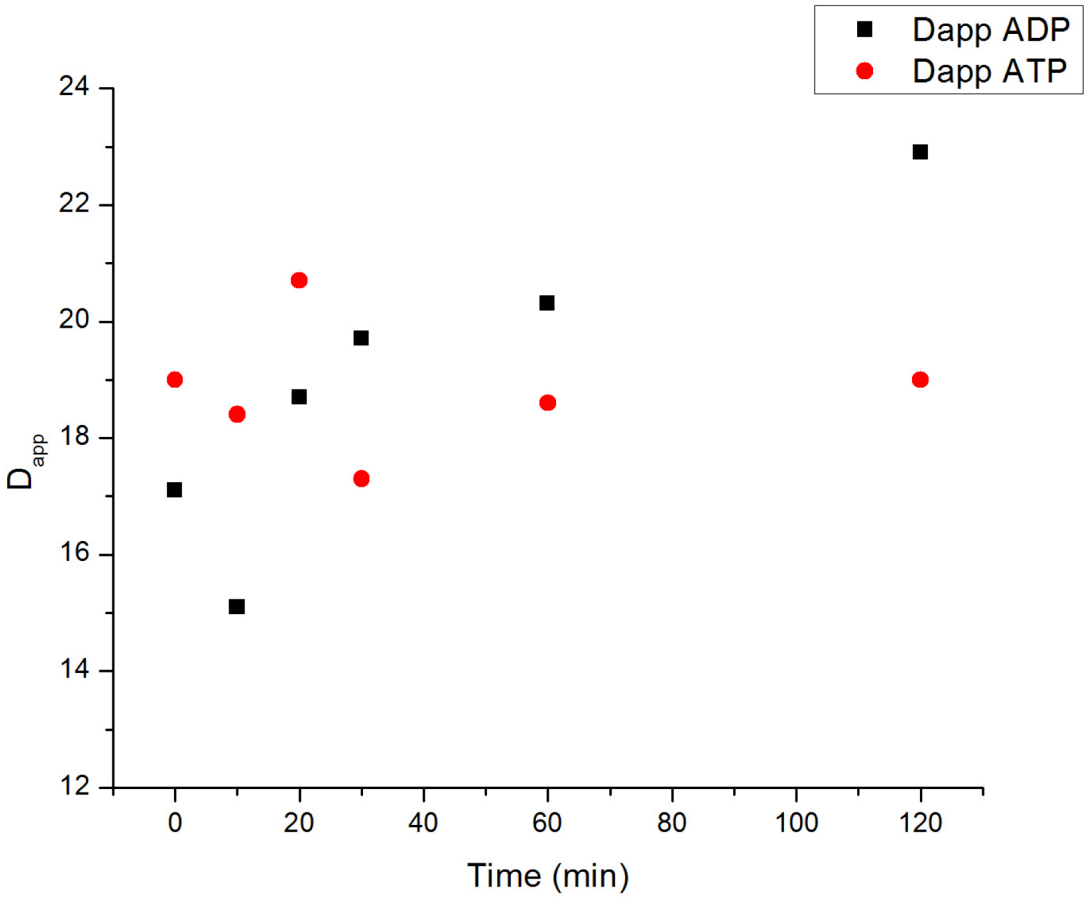


Figure 3.5: Nucleotide incubation comparison in Spinach short-form Rca. D_{app} values collected as a function of time with being incubated in either ADP (black) or ATP (red) while keeping all stock solutions at RT. The concentration of protein used for these experiments was kept constant at 4 µM.

a sample of So-β-Rca incubated in ADP was diluted to 5 µM and filtered using a 0.1 µm pore size spin filter immediately before measurement on the FCS. Doing this filtration step immediately before would in theory retain all misfolded proteins in the filter and allow only smaller oligomers to go through. When normalized and compared to a control sample that was not filtered, it is clear that the aggregates have simply disassembled, passed through the filter, and re-assembled on the other side, suggesting that the large aggregates

are dynamic in nature, not simply misfolded proteins (**Fig. 3.6**). This conclusion is further supported by the same experiment performed using a sample that was incubated at room temperature. Normalization and comparison with a control sample shows the same overall curve shape, indicating the same behavior of the aggregated material (**Fig. 3.6**). To rule out the possibility that this observation was simply an artifact of the concentration, the sample was diluted by a factor of 10. Unsurprisingly, the same trend was observed between the

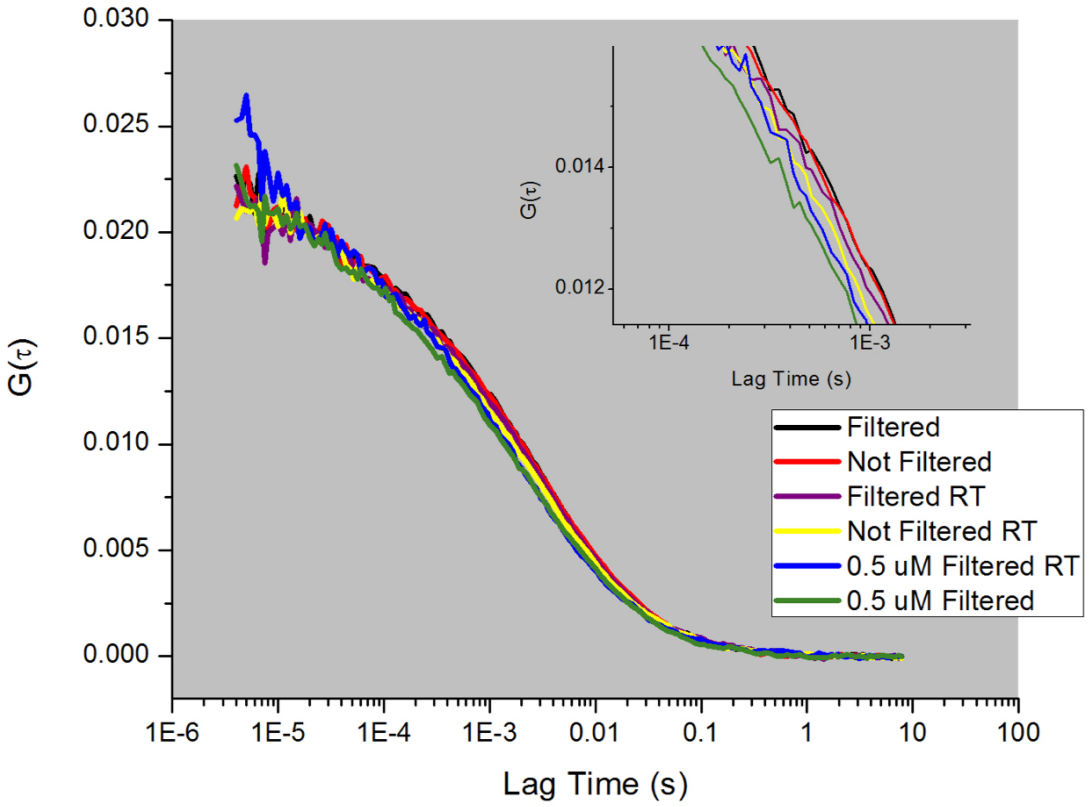


Figure 3.6: **FCS experiments on samples pre and post filtration.** Raw autocorrelation curves collected on Spinach short-form Rca at a concentration of 5 µM that was either filtered (black trace), not filtered (red trace), kept at room temp (RT) and filtered (purple trace), kept at RT and not filtered (yellow trace). The sample was also diluted down 10-fold to 0.5 µM and either filtered (green trace) or not filtered (blue trace). The inset is a zoom in of the traces showing minor differences between the different conditions.

sample filtered immediately before measurement and the one that was not, further supporting the idea that these large aggregates are dynamic not static (**Fig. 3.6**). In addition,

experiments were carried out on material that had never been flash frozen but this did not help with aggregation.

Unique protein induced photo-physics behavior with increasing laser power. In addition to all of the aggregation and incubation issues discussed above, the raw FCS

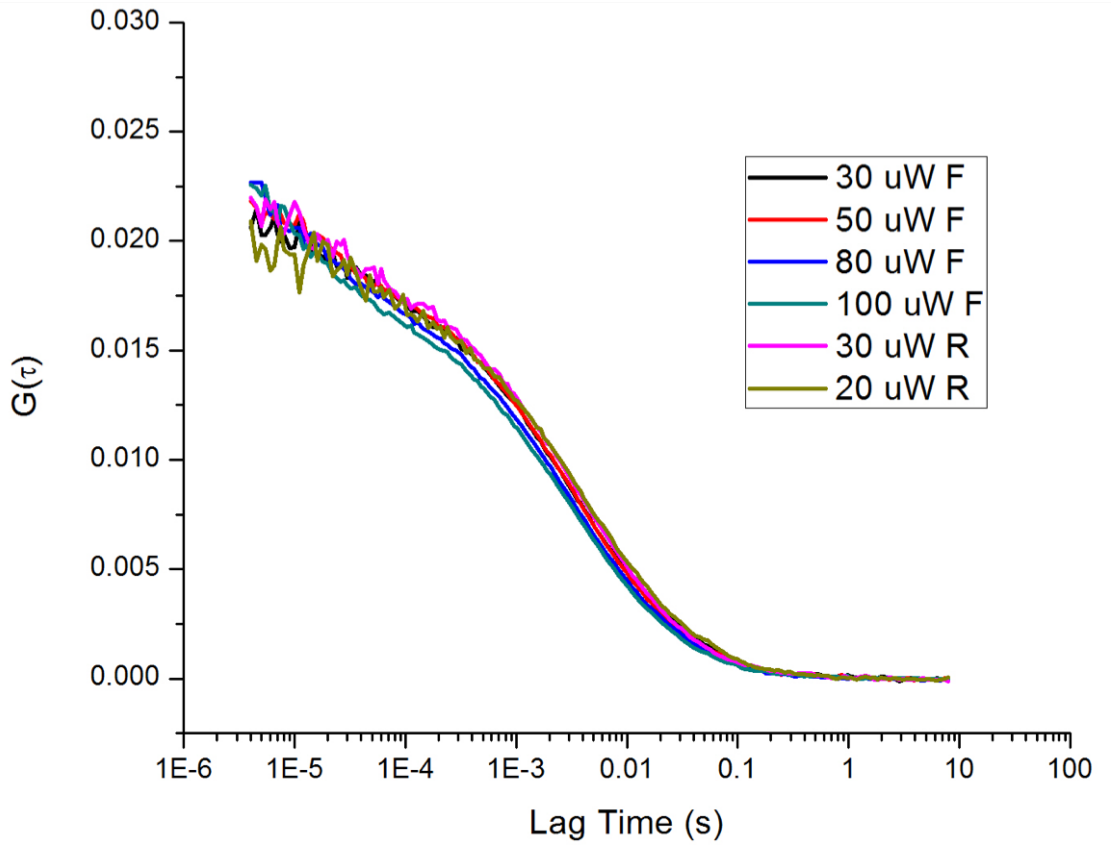


Figure 3.7: Unique photo-physics observed in the autocorrelation curves with changing laser power. Raw autocorrelation curves collected on Spinach short-form Rca at a concentration of 5 μM subjected to different laser powers intensities and normalized with respect to each other. The F and R nomenclature refer to direction they were collected in, either forward (F) or reverse (R).

autocorrelation decays collected contained interesting properties in the short lag time region and varied in response to increasing and decreasing laser power. In order to better understand this behavior, we developed a systematic approach to analyzing the changes in

decay curves with changing laser power. In brief, the So β Rca sample was diluted down to 5 µM with buffer containing ADP as the sole nucleotide and incubated at room temperature for 10 minutes. The sample was then placed on the FCS instrument and

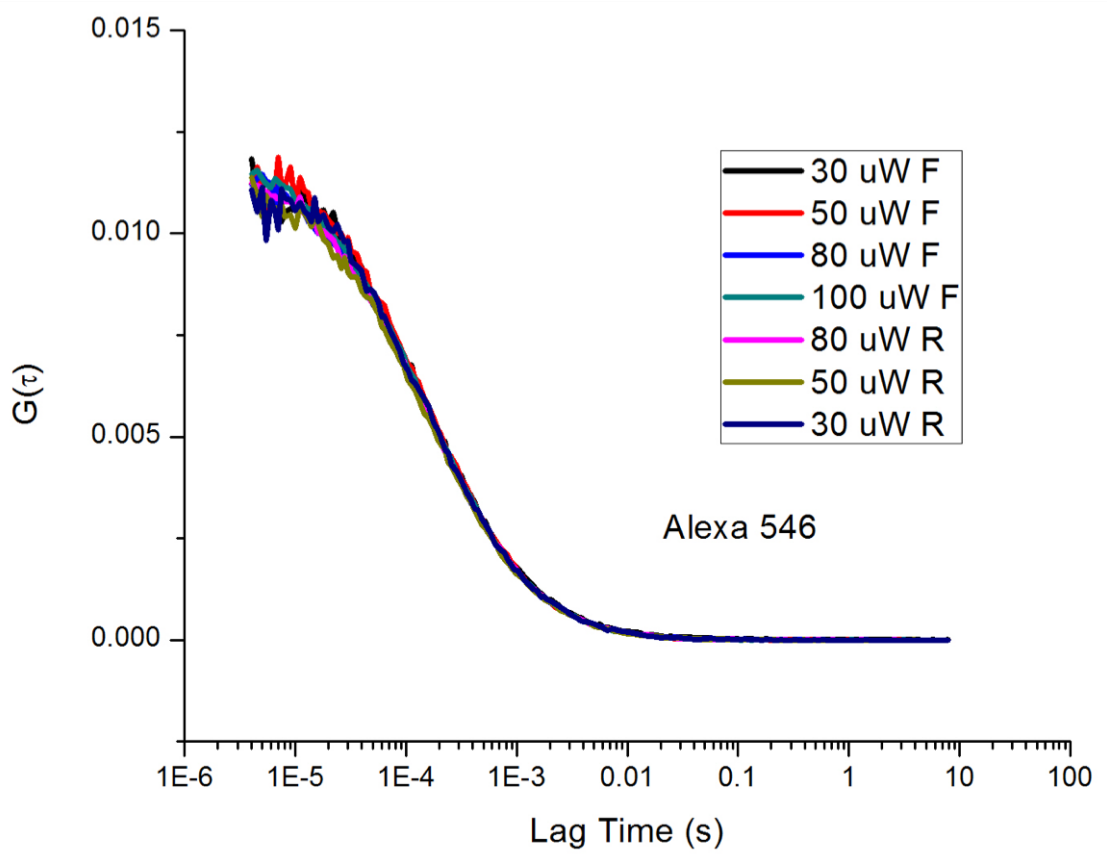


Figure 3.8: Laser power control experiment using an AlexaFluor dye. Raw autocorrelation decays of free AlexaFluor 546 subjected to measurements under different laser powers intensities. Decay curves were normalized with respect to each other for comparative purposes. The F and R nomenclature refer to direction they were collected in, either forward (F) or reverse (R).

measured using a laser power of 30 µW. The same steps were repeated using increasing laser powers, 50 µW, 80 µW, and 100 µW in the forward direction (labeled with F to indicate forward direction), then subsequent decreasing laser powers, 30 µW and 20 µW, in the reverse direction (labeled with R to indicate reverse direction). In contrast to typical

autocorrelation decay curves, decay curves at laser powers greater than 20 μW appear to have a non-flat region at lag times shorter than 1 x 10^{-3} seconds when normalized to together (**Fig. 3.7**). At a laser power of 20 μW, a more typical plateau is observed, suggesting that the increase in laser power affects the fluorescence signal emitted by the labeled sample. This unique feature is certainly attributed to the protein, as a control with free Alexa Fluor 546 shows not such feature in the short lag time region (**Fig. 3.8**).

In order to rule out the possibility of triplet formation, a known possibility for certain dyes with high laser power excitation [36], we added cyclooctatetraene (COT), a known triplet quencher [37] in high excess which ensured no triplet would form. When compared to a sample lacking 20 mM COT, it was clear that this upward trend was not a triplet component given that the sample with quencher did not look different than the sample without the quencher (**Fig. 3.9**). Therefore this unique shape in the autocorrelation decay is not due to the presence of a triplet, but rather seems to be a unique signature of Spinach Rca.

Discussion:

A mixture of nucleotides may be necessary for effectively disassembling large oligomers. Despite the many issue associated with the FCS work, there are still particular trends that can be teased out of the data. One such trend are the differences in assembly when in the presence of different nucleotides. One of the most obvious effects of ATP-γS is that it severely impairs Rca's ability to completely disassemble into smaller oligomers. This trend is supported by previous work showing very poor subunit exchange dynamics when in the presence of this nucleotide [34].

In contrast, FCS work in the presence of ADP as the sole nucleotide appears to

promote disassembly into smaller oligomers, however, does so at a slower rate than in ATP. On the surface, this would appear to be in contradiction with previously published work, however, when examined closer, it may be complimentary. In their work, Peterson-Forbrook *et. al.* observed that addition of ADP facilitated the most rapid response in subunit exchange dynamics [34]. However, their data states nothing about oligomeric state.

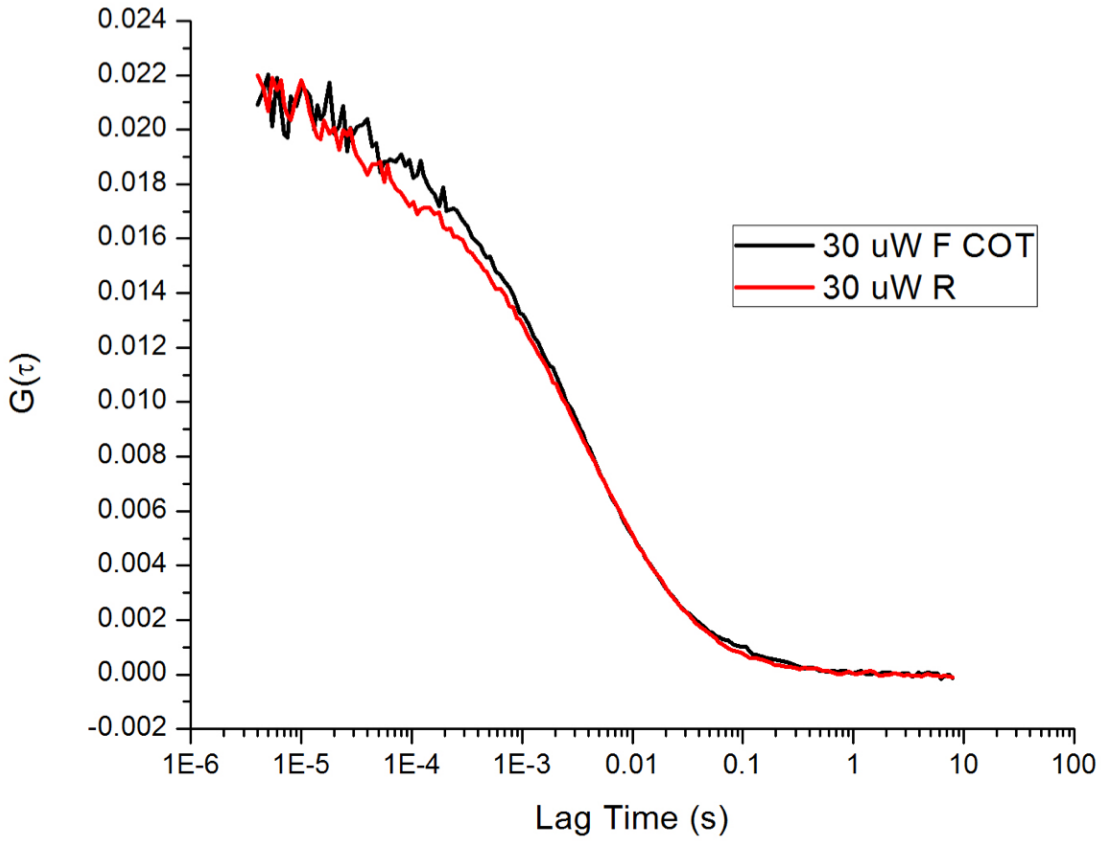


Figure 3.9: Testing COT, a known triplet quencher. Raw autocorrelation decay curves collected on Spinach short-form Rca at a concentration of 5 µM in the presence (black) or absence (red) of 20 mM COT, a known triplet quencher.

Therefore one possibility is that high abundance of ADP does in fact promote the exchange of subunits between large oligomers, but overall disassembly of said large oligomers takes

much longer. From the FCS work on tobacco, it is known that at high protein concentrations, (about 60 μM and above), there is a high abundance of large oligomers and thus it is not a surprise that they would be present when measuring the sample after a short dilution time.

This idea of disassembly is further supported by the measured assembly in the presence of ATP. Unlike, the previous two conditions, ATP can be turned over by Rca to generated ADP, thus generating a mixed nucleotide ratio of ATP to ADP. As was seen with tobacco, this ratio plays a key role in regulating the relative abundance of different oligomers. Therefore the observation that short incubation times in the presence of ATP yield overall smaller oligomers is consistent with the findings from tobacco. It is worth noting that the Peterson-Forbrook *et. al*. found that subunit exchange in the presence of ATP is high, however, it is only high after it accumulates enough ADP to yield a ATP/ADP ratio or roughly 4 [34]. The lag time associated with generating this amount of ADP was determined to be about 5 minutes, which falls well within the incubation time used for FCS measurements.

The alpha-isoform may be necessary to minimize aggregate formation. As was mentioned above, a great deal or work was done to try and eliminate, or a least minimize, the abundance of large aggregates. It is clear that changing the nucleotide or the incubation time does not help with this issues and thus suggests that this propensity to aggregate is a feature specific to the beta-isoform itself, not an issue with any conditions. Additionally, we know that these large aggregates are dynamic in nature, meaning they can quickly assemble and disassemble to pass through a 0.1 μm filter. Given this behavior, the solution would perhaps be to incorporate the longer alpha isoform along with the shorter beta

isoform and repeat the work. This idea comes from work out of the Pearce lab which found that sizing experiments on the alpha isoform carried out using Analytical Ultra Centrifugation (AUC) supported the stabilization of hexamer-like assemblies, while similar experiments using the shorter beta isoform showed larger material [32]. However, when combined at a 1 to 1 ratio, they found that hexamer-like assemblies were still preferred, suggesting that incorporation of the alpha causes formation of proper assemblies. Therefore, incorporation of the alpha isoform may help address and alleviate the aggregate problem. Additionally, it may also address the interesting photo-physical properties observed in the short lag time periods on the FCS measurements.

Conclusion:

In summary, the FCS work presented here does seem to suggest that like tobacco, spinach Rca does alter its assembly in response to varying nucleotides. Given the limitations with aggregation, full assembly curves could not be generated and thus fractional concentrations for various oligomers could not be determined. Despite these hurdles, the fact that the aggregates are dynamic in nature suggests that it is not the quality of the protein that is causing the issues but rather that this is an inherent property of spinach short-form Rca. This further means that improvements can be made to reduce the number of total aggregates, improvements like incorporation of the longer alpha isoform. We hope that the results presented here begin laying the foundation for understanding the underlying factors which govern Rca self-assembly and allow for identification/development of new strategies to address the ever-increasing food and energy shortages our future may bring.

References:

1. Zhu, X-G., *et al.* 2010. Improving Photosynthetic Efficiency for Greater Yield. *Annu. Rev. Plant Biol.* **61,** 235-261

2. Hikosaka, K., and Terashima, I. 1995. A Model of the Acclimation of Photosynthesis in the Leaves of C-3 Plants to Sun and Shade with Respect to Nitrogen use. *Plant Cell Environ.* **18,** 605-618

3. Parry, M. A., *et al*. 2013. Rubisco Activity and Regulation as Target for Crop Improvements. *J. Exp. Biol*. **64,** 717-730

4. Tabita, F. R. *et al*. 2008. Distinct Form I, II, III, And IV Rubisco Proteins From the Three Kingdoms of Life Provide Clues About Rubisco Evolution and Structure/Function Relationships. *J. Exp. Bot.* **59,** 1515-1524

5. Cleland, W.W., *et al.* 1998. Mechanism of Rubisco: The Carbamate as a General Base. *Chem. Rev.* **98,** 549-561

6. Mueller-Cajar, O., *et al.* 2013. Maintaining Photosynthetic CO2 Fixation via Protein Remodeling: The Rubisco Activases. *Photosynth. Res.* **DOI:** 10.1007/s11120-013-9819-0

7. Bauwe, H., et al. 2010. Photorespiration: Players, Partners and Origin. *Trends Plant Sci.* **15,** 330-336

8. Whitney, S.M., *et al.* 2011 Advancing our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco. *Plant Physiol.* **155,** 27-35

9. Tcherkez, G. G. B., *et al*. 2006. Despite Slow Catalysis and Confused Substrate Specificity, all Ribulose Bisphosphate Carboxylases May be Nearly Perfectly Optimized. *Proc. Natl. Acad. Sci*. **103,** 7246-7251

10. Andersson, I. and Backlund, A. 2008. Structure and Function of Rubisco. *Plant Physiol. Biochem.* **46,** 275-291

11. Woodrow, I. E., and Berry, J. A. 1988. Enzymatic Regulation of Photosynthetic CO2 Fixation in C3 Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **39,** 533-594

12. Andersson, I. 2008. Catalysis and Regulation in Rubisco. *J. Exp. Bot*. **59,** 1555-1568

13. Wachter, R. M. and Henderson, J. N. 2015. Rubisco Rescue. *Nature Plants* **1,** 1-2

14. Andralojc, P. J. *et al.* 2012. 2-carboxy-D-arabinitol 1-Phosphate (CA1P) Phosphatase: Evidence for a Wider Role in Plant Rubisco Regulation. *Biochemistry* **442**, 733-742
15. Parry, M. A. J., *et al.* 2008. Rubisco Regulation: A Role for Inhibitors. *J. Exp. Bot*. **59,** 1569-1580

16. Bracher, A., *et al.* 2015. Degradation of Potent Rubisco Inhibitor by Selective Sugar Phosphate. *Nat. Plants* **1,** 14002. DOI: 10.1038/NPLANTS.2014.2

17. Mueller-Cajar, O. 2017. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Front Mol. Biosci.* **4**, 31. DOI: 10.3389/fmolb.2017.00031

18. Portis, Jr. A. R. and Salvucci, M. E. 2002. The Discovery of Rubisco Activase – Yet Another Story of Serendipity. *Photosynth. Res.* **73,** 257-264

19. Mueller-Cajar, O., *et al*. 2011. Structure and Function of the AAA+ Protein Cbbx, a Red-Type Rubisco Activase. *Nature* **479,** 194-199

20. Loganathan, N., *et al.* 2016. Characterization of the Heterooligomeric Red-Type Rubisco Activase from Red Algae. *Proc. Natl. Acad. Sci.* **113,** 14019-14024

21. Bhat, J. Y., *et al.* 2017. Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Mol. Cell* **67,** 1-13

22. Tsai, Y. C., *et al.* 2015. Identification and Characterization of Multiple Rubisco Activases in Chemoautotrophic Bacteria. *Nat. Commun.* **6,** 8883. DOI:10.1038/ncomms9883

23. Sutter, M. *et al.* 2015. Structural Characterization of a Newly Identified Component of α-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Scientific Reports.* **5,** 16243. DOI: 10.1038/srep16243

24. Stotz, M., *et al*. 2011. Structure of Green-type Rubisco activase from Tobacco. *Nat. Struct. Mol. Biol.* **18,** 1366-1370

25. Wendler, P. *et al*. 2012. Structure and Function of the AAA+ Nucleotide Binding Pocket. *Biochim. Biophys. Acta*. **1823,** 2-14

26. Hanson, P. I. and Whiteheart, S. W. AAA+ Proteins: Have Engine, Will Work. *Nat. Rev. Mol. Cell Biol.* **6,** 519-529

27. Portis, Jr. A. R., *et al.* 2008. Regulation of Rubisco activase and its Interaction with Rubisco. *J. Exp. Bot.* **59,** 1597-1604

28. Zhang, N., *et al*. 2001. Characterization of the Regulatory Function of the 46-kD Isoform of Rubisco activase from Arabidopsis. *Photosynth. Res.* **68,** 29-37

29. Chakraborty, M. *et al.* 2012. Protein Oligomerization Monitored by Fluorescence Fluctuation Spectroscopy: Self-Assembly of Rubisco activase. *Biophys. J.* **103,** 949-958

30. Kuriata, A. M. *et al.* 2014. ATP and Magnesium Promote Cotton Short-Form Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Hexamer Formation at Low Micromolar Concentrations. *Biochemistry* **53,** 7232-7246

31. Keown, J. R., *et al.* 2013. Small Oligomers of Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase are Required for Biological Activity. *J. Biol. Chem.* **288,** 20607-20615

32. Keown, J. R. and Pearce, G. 2014. Characterization of Spinach Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Activase Isoforms Reveals Hexameric Assemblies with Increased Thermal Stability. *Biophys. J.* **464,** 413-423

33. Henderson, J. N., *et al.* 2013. Biophysical Characterization of Higher Plant Rubisco Activase. *Biochim. Biophys. Acta*. **1834,** 87-97

34. Peterson-Forbrook, D. S., *et al.* 2017. Nucleotide Dependence of Subunit Rearrangements in Short-Form Rubisco Activase from Spinach. *Biochemistry* **56,** 4906-4921

35. Kurek, I. *et al.* 2007. Enhanced Thermostability of Arabidopsis Rubisco Activase Improves Photosynthesis and Growth Rates under Moderate Heat Stress. *Plant Cell* **19,** 3230-3241

36. Cordes, T., *et al.* 2011. Mechanisms and advancement of antifading agents for fluorescence microscopy and single-molecule spectroscopy. *Phys. Chem. Chem. Phys.* **13**, 6699-6709

37. Ciuba, M. A. and Levitus, M. 2013. Manganese-Induced Triplet Blinking and Photobleaching of Single Molecule Cyanine Dyes. *Chem. Phys. Chem.* **14,** 3495-3502

CHAPTER 4

ATP AND MAGNESIUM PROMOTE COTTON SHORT-FORM RUBISCO ACTIVASE HEXAMER FORMATION AT LOW MICROMOLAR CONCENTRATIONS

Work presented in this chapter reflects previously published work by Kuriata *et al.* As a co-author on this publication, I am including it as part of my thesis, given that I have made significant contributions to this work. Despite not being involved in the planning of the initial experiments, I collected much of the FCS data used to generate Figure 1A of the paper, as at the time, M. Chakraborty and A. Kuriata were both graduating and were busy preparing their dissertations. As mentioned, I collected the third trial of the FCS data for Figure 1A along with additional repeats of the data points which did not have enough trials. After collecting the data, I analyzed it to determine the D_{app} values and submitted it to Dr. J.N. Henderson to do the mathematical modeling of the data. In addition, I collected all last minute data required for getting the manuscript published (verification of $D_{monomer}$ values, normalizing all data with respect to $D_{monomer}$, etc) and edited the final manuscript prior to submission.

The full reference for the article is given below:

Kuriata, A. M. *et al*. 2014. ATP and Magnesium Promote Cotton Short-Form Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Hexamer Formation at Low Micromolar Concentrations. *Biochemistry* **53,** 7232-7246.

Abstract

We report a fluorescence correlation spectroscopy (FCS) study of the assembly pathway of the AAA+ protein Rubisco activase (Rca), a ring-forming ATPase responsible for activation of inhibited Rubisco complexes for biological carbon fixation. A thermodynamic characterization of simultaneously populated oligomeric states appears critical in understanding Rca structure and function. Using cotton beta-Rca, we demonstrate that apparent diffusion coefficients vary as a function of concentration, nucleotide and cation. Using manual fitting procedures, we provide estimates for the equilibrium constants for the step-wise assembly, and find that in the presence of ATPγS, the K_d for hexamerization is ten-fold lower than with ADP (~0.1 vs. ~1 micromolar). Hexamer fractions peak at 30 micromolar and dominate at 8-70 micromolar Rca, where they comprise 60-80% of subunits with ATPγS, compared to just 30-40% with ADP. Dimer fractions peak at 1-4 micromolar Rca, where they comprise 15-18% with ATPγS and 26-28% with ADP. At 30 micromolar Rca, large aggregates begin to form that comprise ~10% of total protein with ATPγS and ~25% with ADP. FCS data collected on the catalytically impaired WalkerB-D173N variant in the presence of ATP provided strong support for these results. Titration with free magnesium ions lead to the disaggregation of larger complexes in favor of hexameric forms, suggesting that a second magnesium binding site with a K_d value of 1-3 mM promotes critical subunit contacts. We propose that closed-ring toroidal hexameric forms are stabilized by binding of Mg·ATP plus Mg^{2+}, whereas Mg·ADP promotes continuous assembly to supra-molecular aggregates such as spirals.

Introduction

Rubisco [ribulose-1,5-bisphosphate (RuBP) carboxylase/oxygenase] is a central enzyme in photosynthesis, as it catalyzes the rate-limiting step in assimilation of inorganic CO_2 into organic building blocks responsible for all the world's biomass (*1*). Implications for agronomic practices, biofuel production and global carbon cycling ensure persistent interest in engineering or selecting for Rubisco enzymes with improved catalytic efficiency, with the ultimate goal of increasing crop yields (*2*). A number of side-reactions undermine Rubisco's carboxylation activity, especially the competing oxygenation of RuBP (ribulose 1,5-bisphosphate) that feeds the unproductive photorespiration pathway. When Rubisco first appeared on the early biotic earth, atmospheric CO_2 levels were many times higher and O_2 much lower than today, so O_2 competition with CO_2 for incorporation into RuBP was not significant (*3*). As eons of oxygenic photosynthesis gradually raised O_2 levels at the expense of CO_2, evolution pushed Rubisco to bind RuBP more tightly, and opened avenues for regulation via non-productive substrate binding and competitive inhibition by phosphorylated metabolites. Hence, a catalytic chaperone, known as Rubisco activase (Rca), came into play to restore activity to inactivated Rubisco complexes by releasing inhibitors through an as yet undetermined mechanism (*4*). Within the chloroplast, the light-dependent stromal energy charge has been proposed to regulate Rca activity via the ATP:ADP ratio (*5*). In species containing the redox-sensitive Rca α-isoform, additional regulation is afforded by the stromal redox potential.

Rca belongs to the AAA+ (ATPase associated with diverse cellular activities) superfamily of ATPases and is part of the extended group (*6*). Despite diversity in sequence and function, a common theme with AAA+ proteins is the use of energy derived from ATP

hydrolysis to perform chemo-mechanical work on substrate macromolecules. AAA+ proteins contain a structurally conserved ATP-binding cassette of approximately 200 - 250 residues consisting of a N-domain harboring the Walker A ($GX_4GK[T/S]$, where X can be any amino acid) and Walker B (Φ_4DE, where the Φs are hydrophobic residues) motifs, attached to a less conserved C-domain typically involved in substrate recognition. A cleft between the two domains provides a pocket for ATP binding, which frequently directs the formation of active multi-subunit arrays (*7, 8*). The stoichiometries of AAA+ subunit assemblies are directly relevant to their biological activity. Oligomerization brings neighboring protomers together, forming a bipartite ATP binding-site with a conserved arginine (arginine finger) from one protomer interacting directly with the γ-phosphate of ATP bound to the adjacent protomer (*8*). Presumably, this arrangement allows catalytic cycles involving ATP binding, hydrolysis and ADP release to facilitate conformational changes that in turn mediate substrate remodeling or translocation events. Some AAA+ proteins require macromolecular binding partners to act as scaffolds for proper assembly (*9*), and often there is a requirement for nucleotide binding (*10-13*). Numerous reports suggest that assemblies with subunits in hexameric arrangements are most common (*13-17*), though there are some exceptions (*18-20*).

The highly polydisperse nature of protein preparations of Rca has long hampered mechanistic studies. Size exclusion chromatographs exhibit broad, asymmetric bands, suggesting concentration-dependent size distributions that range from monomers to extremely large aggregates of nearly 50 subunits (*21-24*). However, recent work has demonstrated that gel filtration does not reflect equilibrium conditions, in line with rapid Rca subunit exchange (*24*). Regardless, nucleotide-dependent Rca oligomerization

appears to be essential for proper biological function (*21, 23, 25, 26*), and recent dynamic light scattering (DLS) experiments have confirmed that ATP, ADP and Mg^{2+} increase self-association of cotton β-Rca (*24*). Polydispersity was also observed by NanoESI mass spectrometry of wild-type tobacco Rca (Nt-β–Rca) in the presence of ADP, where sizes ranged up to hexameric forms (*27*). However, the stable formation of a hexamer could only be demonstrated in the presence of ATPγS for a mutant Rca bearing the single-site substitution R294A.

In 2011, a bacterial red-type and two higher-plant green-type Rca X-ray structures became available, all demonstrating a classic AAA+ architecture (*28-30*). For red-type Rca from *Rhodobacter sphaeroides* (Rs-Rca or CbbX), negative stain electron microscopy (EM) demonstrated the formation of ring-like hexamers in the presence of ATP and RuBP. In the absence of nucleotides, amorphous particles of ~600-900 kDa were observed, whereas Mg·ATP without RuBP provided long fibers of ~5-10 MDa (*28*). Based on a mutational analysis of pore-lining residues, a model was proposed that involves the transient pulling of the Rubisco large subunit C-terminal peptide into the central pore of the Rs-Rca hexameric toroid, thereby facilitating release of inhibitory sugar phosphates.

Higher-plant Rca structures consist of the 1.9 Å resolution *Larrea tridentata* (creosote bush) substrate recognition domain, a helical bundle that forms the core of the AAA+ C-domain (*29*), and the 2.9 Å resolution nucleotide-free *Nicotiana tabacum* (tobacco) core AAA+ module, which crystallized in a hexameric spiral subunit arrangement (*30*). By EM, closed-ring hexamers were observed for the Nt-β–Rca-R294V variant in the presence of Mg·ATP or Mg·ATPγS (adenosine 5'-O-[gamma-thio]triphosphate), whereas Mg·ADP produced amorphous particles (*30*). Mutational studies of central pore residues reduced or

completely abolished Rubisco reactivation activity, again suggesting a role in peptide threading (*30*). Although several Rca - Rubisco binding geometries have been proposed, the details of the physical interaction between these complexes await an explicit description (*31*).

Despite the hexameric arrangements observed in EM images, mixing experiments of wild-type and mutant tobacco Rca led to the conclusion that three to six subunits are necessary for Rubisco activation. These data suggest that hexamerization may not be a strict requirement for function (*30*). Using static light scattering (SLS), analytical ultracentrifugation (AUC) and small-angle X-ray scattering (SAXS) on tobacco Rca, Keown et al. arrived at an estimate of two to four subunits as the minimal requirement for both ATPase and Rubisco reactivation activities (*32*). As in previous works, the apparent mass of apo-Rca was shown to increase continuously with increasing subunit concentration, with no apparent accumulation of particular intermediate forms.

To-date, experiments aimed at elucidating the mechanistic enzymology of Rca continue to be hampered by a lack of understanding of the mix of Rca complexes that coexists in solution under various conditions. Therefore, kinetic constants such as hydrolytic turnover numbers and Rubisco activation rates can only be interpreted as an average activity per Rca subunit (*33*). Recently, we have developed fluorescence correlation spectroscopic (FCS) methods to determine individual binding constants for the step-wise assembly of cotton β-Rca (*34*). FCS relies on the temporal fluctuation in fluorescence intensity that arises when a small number of molecules is present in an optically restricted volume (*35, 36*). Autocorrelation analysis relates these fluctuations to dynamic processes such as Brownian motion. For the first time, application of this method has allowed for the

quantitative determination of the fractional population each Rca oligomer present at specific subunit concentrations in the presence of Mg·ADP (*34*). In this work, the K_d values for Mg·ADP-mediated assembly were estimated to be 3.5 μM for the monomer-dimer equilibrium ($K_{m\text{-}d}$), 1 μM for the dimer-tetramer equilibrium ($K_{d\text{-}t}$), and 1 μM for the tetramer-hexamer equilibrium ($K_{t\text{-}h}$). The FCS results suggest that as much as four different oligomeric states coexist at cotton-β-Rca concentrations typically used in ATPase and Rubisco reactivation kinetic assays (*33*). As self-association did not provide any plateaus under these conditions, the continued assembly to higher order states was modeled as a hexamer-24mer equilibrium ($K_{h\text{-}24}$) with a K_d value of 25 μM^3. Therefore, we have interpreted the ADP-mediated mechanism of assembly in terms of continued growth along a helical axis that is defined by an open spiral assembly (*34*), in which each turn consists of 6 protomers as observed in the tobacco Rca crystal structure (*30*). We have proposed that this mechanism is similar but not equal to the physiological assembly pathway, which likely involves closed-ring hexameric toroids (*34*).

In the present work, we have applied the same FCS method to the investigation of cotton β-Rca assembly as a function of ligands that more closely mimic physiologically relevant conditions. To address the effects of different ATP levels in the chloroplast stroma, we have utilized the slowly hydrolyzing ATP analog ATPγS in oligomerization assays. We estimate that close to 80% hexamer can be generated *in vitro* over a narrow Rca concentration range. To confirm these results, we have investigated the early assembly steps of a substrate-trap mutant as a function of different ATP:ADP ratios. Notably, the data indicate the presence of a second Mg^{2+} binding site on the protein (in addition to the

primary ATP-based coordination site) that appears to promote the formation of smaller oligomers at the expense of supra-molecular aggregates.

Materials and Methods

Site-directed Mutagenesis: A modified version of the QuickChange method (Stratagene, Lo Jolla, CA) was used to introduce the D173N substitution into a *Gossypium hirsutum* (cotton) β-Rca harboring a C-terminal Ala-Cys insertion (Gh-β-Rca-378AC). This construct, termed β-Rca-AC throughout the text, was previously cloned into a pET151-dTOPO expression plasmid (*34*). The following forward and reverse oligonucleotide primers were ordered from Integrated DNA Technologies (Coralville, IA): 5'-CTCTTCATCAACAATCTCGACGCTGGAG-3' and 5'-CTCCAGCGTCGAGATTGTTGATGAAGAG-3'. PfuTurbo DNA polymerase and 10X Pfu DNA polymerase reaction buffer were purchased from Agilent Technologies (Santa Clara, CA), dNTPs from Life Technologies (Grand Island, NY) and DpnI restriction enzyme from New England Biolabs (Ipswich, MA). Thermal cycling, DpnI digestion and heat shock transformation into chemically competent Top10 *E.coli* cells was carried out according to the QuickChange Site-directed Mutagenesis guidelines. Single colony transformants from overnight LB (Luria-Bertani) agar plates containing 100 μg/mL carbenicillin were cultured in 25 mL of LB with 100 μg/mL carbenicillin, and grown overnight at 37°C and 250 rpm. After centrifugation (5000 rpm, 4°C, 10 minutes), plasmids were purified from overnight cultures using the centrifugal method from the Qiagen Plasmid Mini Prep Kit (Germantown, MD). The correct mutation was confirmed by DNA sequencing.

Expression, purification and dye conjugation of Rca: β–Rca-AC and its variants were recombinantly expressed from pET151-dTOPO or pET23 plasmids, with and without N-terminal 6XHis affinity tags respectively, as previously described (*34*). β-Rca-AC expressed without an affinity tag was purified according to methods published earlier (*24*). Nucleotides and other reagents used were of the highest purity reasonably obtainable. Frozen cell pellets (-80°C) from 3 L of culture expressing β-Rca-AC or β-Rca-AC-D173N were thawed on ice and suspended in 50 mL of 25 mM Tris-HCl pH 8.0, 10 mM imidazole pH 8.0, 10% glycerol, ~30 μM hen egg white lysozyme (Sigma-Aldrich), 1 mM phenylmethanesulfonylfluoride, 1 mM DTT, 0.1 mM ADP and 0.1 mM EDTA; then stirred at 4°C for 30 minutes and disrupted by sonication. Cell lysate was pelleted by centrifugation, and the supernatant passed through a 0.8 μm syringe filter before being loaded onto a Ni^{2+}-nitrilotriacetic acid (Ni-NTA) column (Qiagen, Valencia, CA). Protein was purified using an imidazole buffer step gradient (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT and 0.1 mM ADP, plus variable amounts of imidazole). After washing the columns with 10 mM and 80 mM imidazole buffer, His-tagged β-Rca-AC and β-Rca-AC-D173N were eluted from the column with 35 mL of 200 mM imidazole buffer. The pooled elution fractions were digested with tobacco etch virus (rTEV) protease to remove the 6XHis tag, dialyzed overnight, passed over a second Ni-NTA column and concentrated to 2.5 mL in Centriprep concentrators (Millipore, Milford, MA) as described (*34*). Samples were then buffer exchanged into 25 mM HEPES-NaOH pH 7.5, 250 mM KCl, 2 mM nucleotide (ADP, ATP, ATPγS or AMP-PNP [5'-adenylyl-imidodiphosphate]), 5 mM $MgCl_2$ (or no $MgCl_2$) and 10% glycerol using PD-10 columns (GE Healthcare). Protein

concentrations were determined using the Bradford method, with typical yields of 3-4 mg/L cell culture and 1.6 mg/L cell culture for β-Rca-AC and β-Rca-AC-D173N, respectively. β-Rca-AC and β-Rca-AC-D173N conjugation with an Alexa-546 C_5-maleimide fluorophore and analysis for efficiency of dye labeling followed the published method (*34*), except that ALEXA dye stock was prepared at 350 μM in 50 mM HEPES-NaOH pH 7.2, rather than at 1.93 mM in 50 mM Na_2HPO_4/NaH_2PO_4 pH 7.2.

HPLC, spectrophotometric and mass spectrometric analysis of labeled protein: Alexa-labeled β-Rca-AC and β-Rca-AC-D173N samples were analyzed by reverse-phase HPLC (Agilent Technologies 1260 Infinity Quaternary LC system, Agilent Technologies 1100 Series Diode-Array detector) on a C18 analytical column (Phenomenex Prodigy, 5 μm ODS-3, 100 Å pore size, 250 x 4.6 mm) using a linear water/acetonitrile gradient with 0.1% trifluoroacetic acid (TFA). Protein was monitored by optical density (O.D.) at 220 nm and 280 nm, and Alexa dye by O.D. 550 nm. Whereas free Alexa eluted at 40.5 min, all protein eluted at 36.5 min . Protein fractions were collected by hand, and absorbance spectra measured using 50% acetonitrile, 0.1% TFA as a blank. From these spectra, molar ratios for Alexa/protein were calculated using the previously described method (*34*). To verify correct molecular mass of labeled samples, MALDI spectra were collected on a Voyager DE STR mass spectrometer as previously described (*34*).

General FCS sample preparation method: In typical FCS experiments, 500 μL of FCS sample buffer (25 mM HEPES pH 7.6, 250 mM KCl, 2 mM nucleotide or nucleotide analog, 5 mM $MgCl_2$, and 10% glycerol) was placed on ice and used for sample dilutions. Protein preparations were buffer-exchanged into FCS sample buffer at Rca subunit concentrations of 85-90 μM for labeled and 110-135 μM for unlabeled protein.

Subsequently, aliquots were flash-frozen with liquid nitrogen and stored at -80°C. On the day of the FCS experiment, aliquots were thawed on ice, and both 1 μM and 10 μM stocks of unlabeled protein were prepared, to be used in subsequent dilutions of labeled protein. All aliquots were kept on ice until measurement. For each experiment, appropriate amounts of labeled and unlabeled protein were mixed in siliconized low-retention microcentrifuge tubes containing 1X FCS sample buffer to give the desired final protein concentration. Regardless of total protein concentration, all samples were prepared so as to contain exactly 50 nM labeled protein. To allow subunit equilibration, samples ready for analysis were incubated for approximately 15 minutes on ice before performing FCS experiments.

For concentration-dependent experiments at different nucleotide ratios, aliquots of labeled and unlabeled β-Rca-AC-D173N containing either ATP or ADP were first mixed to give the desired nucleotide ratio, followed by dilution to the appropriate subunit concentration for the FCS measurement. For these experiments, the nucleotide content of the FCS buffers was adjusted to contain either 2.0 mM ATP, or 1.5 mM ATP plus 0.5 mM ADP, or 2 mM ADP. For low $MgCl_2$ experiments, buffer contained 1.5 mM ATP, 0.5 mM ADP and 2.02 mM $MgCl_2$.

FCS sample preparation for Mg^{2+} titration experiments at fixed protein concentrations: A set of four FCS sample buffer solutions were prepared, each containing 1.5 mM ATP and 0.5 mM ADP with a differing total $MgCl_2$ concentration (1.7, 3, 5 or 7 mM). Samples for FCS containing 45 μM β-Rca-AC-D173N were prepared by mixing labeled protein with unlabeled protein following a dilution protocol analogous to the one described above. The concentration of free Mg^{2+} in solution was calculated assuming multiple equilibria and

using the reported dissociation constants for ADP and ATP of 0.241 mM and 0.028 mM, respectively. Because the amount of ADP and ATP was fixed throughout the experiment, this allowed the concentration of free Mg^{2+} in the final sample to be adjusted simply by diluting with the appropriate amounts of one of the four FCS sample buffers with variable $MgCl_2$.

FCS sample preparation for nucleotide titration experiments at fixed protein concentration: For this experiment, labeled and unlabeled protein stock solutions were prepared by mixing appropriate amount of thawed β-Rca-AC-D173N in ATP or in ADP, following dilution protocols as described above. Each experiment contained a total of 60 μM β-Rca-AC-D173N, of which 50 nM was labeled. FCS sample buffers with differing nucleotide compositions were used to dilute protein samples to the desired final concentration of 60 μM. Nucleotide composition was 2 mM ATP, 1.5 mM ATP plus 0.5 mM ADP, 1 mM ATP plus 1 mM ADP, 0.5 mM ATP plus 1.5 mM ADP, and 2 mM ADP.

Fluorescence fluctuation spectroscopy data collection: FCS measurements were carried out using a home-built confocal optical setup. Excitation was achieved with a Compass 215M-10 532 nm CW laser (Coherent GmbH, Germany) attenuated to 50 μW to minimize triplet dynamics. The output of the laser was expanded, collimated, and directed via a dichroic lens into an Olympus PlanApo 100X/1.4NA Oil objective. Samples were placed into silicone perfusion chambers (Grace Biolabs, Bend, OR) pre-treated with BSA to minimize Rca adsorption onto the cover glass. Fluorescence from samples was collected via the same objective, separated from excitation light through the dichroic, and reflected into a 50 μm pinhole. The emission was then focused into an avalanche photodiode detector (Perkin-Elmer Optoelectronics, SPCM-AQR14). A bandpass filter was used

before the detector to minimize stray light (Omega 3RD560-620). The recorded fluorescence signal was autocorrelated in real time using an ALV5000/EPP USB-25 correlator (ALV GmbH, Germany).

FCS data analysis: The autocorrelation function of a single species that diffuses freely in solution is given by (*35*):

$$G(\tau) = G_0\left(1+\frac{4D\tau}{r_0^2}\right)^{-1}\left(1+\frac{4D\tau}{z_0^2}\right)^{-1/2}$$

where τ is the correlation lag-time, G_0 is the amplitude of the decay, D is the diffusion coefficient of the diffusing particle, and r_0 and z_0 are the radial and axial semi-axes of the Gaussian confocal volume.

The instrument was calibrated daily by measuring the FCS decays of free TAMRA dye (D = 414 $\mu m^2 s^{-1}$) in buffer (*37, 38*), and fitting the experimental decays using r_0 and z_0 as fitting parameters. The diffusion coefficient of Rca was determined by fitting the experimental FCS decays using the values of r_0 and z_0 measured on the same day. As shown in our previous work, solutions of Rca are in general polydisperse and therefore the diffusion coefficients obtained from the fits represent apparent values for a given mixture of oligomeric species. Data are presented as normalized ratios D_{app}/D_1, where D_1 is the diffusion coefficient of the monomer measured with a 50 nM Rca subunit concentration.

General parameters of the Enzcheck phosphate assay: Inorganic phosphate (Pi) production was monitored with a UV 2401 PC Shimadzu spectrophotometer using the Enzcheck phosphate assay (Life Technologies) (*39*). This assay is a continuous, enzyme-linked assay in which the substrate 2-amino-6-mercapto-7-methyl purine riboside (MESG, λ_{max} = 330 nm) is converted to ribose-1-phosphate and 2-amino-6-mercapto-7-methyl

purine (λ_{max} = 360 nm) by purine nucleoside phosphorylase (PNP). The coupling enzyme PNP consumes phosphate during turnover, such that ATPase activity is reflected by an increase in absorbance at 360 nm.

The standard curve for the Enzcheck assay was prepared in the following manner. Reference and sample cells containing 1X Enzcheck reaction buffer (50 mM Tris-HCl pH 7.5, 1 mM $MgCl_2$ and 0.1 mM NaN_3) supplemented with 55 mM HEPES pH 8, 1.6 mM DTT and 4 mM $MgCl_2$ were auto-zeroed at 360 nm. All concentrations provided refer to the final volume of 1.00 ml in the cuvette. MESG and PNP were added sequentially to the sample cuvette to provide a final concentration of 200 μM and 3 U/mL, respectively. The mixture was allowed to equilibrate for 30 seconds before starting the time scan. Once the absorbance signal stabilized, known concentrations of phosphate (P_i) standard (KH_2PO_4) were added, manually mixed with a cuvette mixer, and time scans were continued until signal saturation was reached. The final reaction volume was 1 mL for each measurement, and is reflected in all concentrations given above. For each P_i concentration, initial and final O.D.360 values were recorded, and the difference plotted as a function of [P_i]. A linear equation fit to these data represented the standard curve that was subsequently used to calculate the rate of P_i release upon Rca-catalyzed ATP hydrolysis.

Measurement of ATPase activity of β-Rca-AC and β-Rca-AC-D173N using the Enzcheck assay: The ATPase activity of 5 μM Rca was assayed in buffer containing 4 mM ATP, 60 μM ADP and 5 mM $MgCl_2$, and also in buffer containing 0.5 mM ATP, 60 μM ADP and 5 mM $MgCl_2$. Each condition was measured in triplicate. For this assay, substrate ATP was prepared in 100 mM HEPES pH 8.0 to avoid acid hydrolysis. Reference and sample cells were auto-zeroed as described above for the P_i standard curve. Next, 200

nmol MESG, 5 nmol Rca and 3 Units PNP were added sequentially, and the mixture was equilibrated for 1 minute to consume contaminating phosphate until a stable signal was obtained. The time scan was started and, after 1 minute, substrate ATP (4 μmol) was added, the sample mixed manually, and the time scan continued for another 4 minutes. P_i contamination from ATP gives a significant signal even in the absence of Rca. Therefore, a blank experiment was performed, in which a buffer blank (25 mM HEPES, 300 mM KCl, 1 mM DTT, 2 mM ADP and 10% glycerol) was added instead of Rca. A second control experiment consisted of the addition of 5 μM bovine serum albumin (BSA, Sigma Aldrich) instead of Rca, with BSA suspended in 25 mM HEPES, 300 mM KCl, 1 mM DTT, 2 mM ADP and 10% glycerol. To calculate initial rates, the time scan collected on a buffer blank was subtracted from Rca or BSA turnover scans. The data range chosen for initial rate calculation consisted of the steepest linear portion of the time scan, which was generally 20-30 seconds long. The corrected initial slope was converted to the rate of phosphate (P_i) production by use of the standard curve. The turnover rate of Rca was calculated by dividing the rate of change in phosphate concentration by the subunit concentration of Rca.

Measurement of ATP hydrolysis during β-Rca-AC-D173N preparation for FCS experiments. The phosphate content of β-Rca-AC-D173N was determined under conditions identical to those used for FCS sample preparation. Frozen (-80°C) aliquots of 113 μM protein in 25 mM HEPES-NaOH pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 2 mM ATP were thawed on ice, alongside corresponding buffer samples. 22.1 μL of protein solution was transferred to a cuvette containing 0.2 mM MESG, 3 Units of PNP, 50 mM Tris pH 8.0, 55 mM HEPES pH 8.0, 1 mM $MgCl_2$, and 0.1 mM NaN_3, to provide a final volume of 1 mL in the cuvette. After mixing, the signal was allowed to

stabilize, and the O.D.360 was recorded. Buffer blanks were utilized for baseline correction, and the corrected O.D.360 was utilized to determine the phosphate content in the sample using an appropriate standard curve (see above). Additional time points were collected upon protein incubation for 20, 175 and 300 min on ice. The phosphate content of the protein sample calculated from these data was assumed to result from ATP hydrolysis and therefore to be equivalent to the ADP content. Traces of phosphate arising from sources other than ATP hydrolysis were taken into account by the baseline correction method.

Measurement of ATPγS hydrolysis during β-Rca-AC preparation for FCS experiments. A number of protein samples used for FCS contained the slowly hydrolyzing nucleotide analog ATPγS. To quantitate the extent of ATPγS hydrolysis during sample preparation for FCS experiments, a nicotinamide adenine dinucleotide (NADH) enzyme-linked assay was used to measure the ADP content (*40*). Frozen (-80°C) samples containing 5 mg/mL protein in 25 mM HEPES-NaOH pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 2 mM ATPγS were thawed on ice alongside corresponding buffer samples. 25.5 μL of protein solution was added to a cuvette containing 0.3 mM NADH, 2 mM phoshoenolpyruvate, 100 mM tricine pH 8.0, 5 mM $MgCl_2$, 20 mM KCl, 4.8-8 U pyruvate kinase and 7.2-11.2 U lactate dehydrogenase, to provide a final volume of 1 mL. Blank measurements were recorded upon addition of 25.5 μL buffer. The signal at 340 nm was observed to decrease with time, then to stabilize to a constant value. The difference between the initial and final O.D.340 readings, corrected by measurements on the buffer blank, was converted to the amount of NADH consumed by use of the NADH extinction coefficient (ε_{340}) of 6,220 $M^{-1}cm^{-1}$ (*41*). This value was set equal to the amount of ADP

produced by ATP hydrolysis. The measurement was repeated for ice incubation times of 20 minutes, 175 minutes and 300 minutes.

Thermofluor stability assays. Fluorescence data were collected with an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). β-Rca-AC and β-Rca-AC-D173N premixes at 1.05X were prepared as described previously (*24*), and 19 μL were pipetted into the wells of 384-well polypropylene TempPlate PCR plate, which was kept on ice. Nucleotides (ADP, ATP, ATPγS or AMP-PNP) at 2 mM, $MgCl_2$ at 5 mM, or both were introduced by addition of 1 μL stocks containing either 40 mM nucleotide, 100 mM $MgCl_2$ or both. For lower nucleotide or magnesium content, buffers were adjusted accordingly. Each individual well of the TempPlate PCR plate contained a final volume of 20 μL sample containing 0.25 mg/mL Rca, 25 mM HEPES-NaOH pH 7.5, 150 mM KCl, 16X Sypro Orange (Invitrogen), and appropriate amounts of nucleotide and magnesium ions. Sypro Orange is supplied as a 5000X solution by the manufacturer. Each buffer condition was set up in triplicate. The plate was sealed with optically clear tape, and data were collected and processed as previously described (*24*).

Results

ATPγS promotes hexamerization of cotton β-Rca-AC, while ADP favors larger oligomers. To better understand nucleotide-dependence of Rca subunit assembly, we have used previously developed FCS methods to monitor oligomerization in the presence of the slowly hydrolyzing ATP analog ATPγS, and compared the results with the assembly process in the presence of ADP (*34*). As described previously, a fluorescent label was attached to the Rca C-terminus by employing a carboxy-terminal Ala-Cys insertion variant

originally derived from *Gossypium hirsutum* (cotton) short-form (β) Rca (*42*). Here, this variant is termed β-Rca-AC and is also referred to as wild-type protein (loosely defined). This protein was expressed in *E. coli* as an N-terminally 6His-tagged fusion protein and purified by Ni-affinity chromatography. The tag was cleaved with rTEV (recombinant tobacco etch virus) protease, and the thiol-directed label AlexaFluor-546 was attached to the protein's engineered cysteine residue by means of maleimide chemistry (*34*). We were able to efficiently incorporate one label per protomer for all reactions carried out in the presence of ATPγS, as previously demonstrated for ADP-containing samples (*34*). Absorbance spectra collected on HPLC-purified and acid-denatured protein were consistent with a molar ratio of 1.1 (n = 1) labels per protein chain, as further supported by MALDI mass spectrometry (see Supplementary Figures S1 and S2). Enzymatic assays monitoring ADP production were employed to determine the amount ATPγS hydrolyzed by β-Rca-AC under conditions mimicking sample preparation on the day of the FCS experiment (typically 5 hours for the measurement of a full Rca concentration range) (Table S1). These data were utilized to calculate an average ATPγS:ADP ratio of 8 (range 7 – 11) to be present in Rca· ATPγS samples during FCS data collection (Figure 1A).

Unfortunately, similar experiments could not be carried out in the presence of the non-hydrolyzable ATP-analog AMP-PNP (**β,γ-imidoadenosine 5′-triphosphate)**, because Rca·AMP-PNP MALDI spectra provided evidence for the incorporation of two labels per protein chain, suggesting that a normally buried native cysteine residue has become solvent-accessible. Thermal destabilization of β-Rca-AC due to AMP-PNP binding was

further supported by Thermofluor assays as described below (Table 1). Therefore, FCS analysis of β-Rca-AC in the presence of AMP-PNP was not pursued further.

β-Rca-AC samples were prepared between 50 nM and 100 μM subunit concentration, with each sample containing 50 nM dye-conjugated protein, 2 mM ATPγS and 5 mM total $MgCl_2$ at pH 7.6. Control experiments indicated that mixing of labeled and unlabeled Rca preparations provided complete subunit equilibration within 10 min incubation on ice, in line with a random distribution of labeled and unlabeled subunits. Under these conditions, the free magnesium ion concentration, i.e. Mg^{2+} not coordinated to ATPγS, was estimated to be 3.02 mM based on its ATP-binding constant (*43*). This compares to a free Mg^{2+} concentration of 3.14 mM in the presence of 2 mM ADP under otherwise identical conditions.

As described previously (*34*), the autocorrelation function was utilized to calculate FCS decay curves from the fluorescence intensity fluctuations (Supplementary Figure S3). These curves were fit to an apparent diffusion coefficient (D_{app}) that contains contributions from all oligomeric species present in the observation volume during measurement. At Rca subunit concentrations between 50 and 300 nM, the FCS decays of β-Rca-AC with Mg·ATPγS provide a relatively constant D_{app} (Figure 1A). Photon counting histogram (PCH) analysis confirmed that β-Rca-AC is monomeric over this concentration range, as demonstrated previously for Mg·ADP bound samples (*34*). When the subunit concentration was raised above 300 nM, the FCS decay curves shifted to longer timescales, indicating Rca oligomerization (Supplementary Figure S3). Figure 1 presents a plot of the D_{app}/D_1 values ($D_1 = D_{monomer}$) as a function of total subunit concentration in the presence

of Mg·ATPγS (Panel A), as compared to previously published data with Mg·ADP (*34*) (Panel B). Although the two curves are very similar from 50 to 700 nM Rca, a steeper decrease in D_{app}/D_1 is observed between 0.7 and 15 μM with Mg·ATPγS than with Mg·ADP. This observation suggests that in the low-micromolar range, oligomer formation is promoted by ATPγS. Strikingly, the opposite trend is observed at concentrations ranging from 30 to 100 μM Rca, where continued assembly to higher-order aggregates appears to be attenuated by Mg·ATPγS compared to Mg·ADP (Figure 1).

To provide a theoretical scale for the expected diffusion coefficients for monodisperse β-Rca-AC, D_{app}/D_1 values were calculated from the molecular weight of pure monomeric (1), dimeric (2), trimeric (3), etc. species by assuming a spherical shape (Figure 1, blue horizontal bars). As shown previously (*34*), utilizing the radius of gyration of available X-ray models (*30*) instead of a spherical approximation did not modify the calculated diffusion coefficients substantially. To characterize the thermodynamics of Rca assembly, we employed a previously developed mathematical model that describes the total autocorrelation function in terms of individual contributions from each oligomeric species (*34*). This analysis allows for the input of K_d values as variable parameters to simulate expected D_{app} values as a function of Rca concentration. As before, two assembly mechanisms were tested (*34*). The first involved the concerted assembly of six monomers to form a hexamer, followed by association of four hexamers to form a 24-mer. The second involved the step-wise assembly of monomers to dimers, dimers to tetramers, tetramers to hexamers, and hexamers to 24-mers, thus requiring the input of four K_d values. In both models, the assembly state of 24 subunits was used as a placeholder for higher-order aggregates, as the FCS data cannot distinguish large species that differ by a few subunits

(Figure 1). For each model, iterative rounds of K_d optimizations were carried out by systematically varying the input K_d values both one at a time and simultaneously, to arrive at predicted D_{app}/D_1 values that follow the experimental data closely.

As observed previously for Mg·ADP-mediated assembly (*34*), the Mg·ATPγS data do not fit well to the concerted model. Therefore, species intermediate to monomers and hexamers must be explicitly modeled (Figure 1A, red curve). A good fit to the data is obtained with $K_{m\text{-}d}$ = 6 μM for the monomer-dimer equilibrium, $K_{d\text{-}t}$ = 2 μM for the dimer-tetramer equilibrium, $K_{t\text{-}h}$ = 0.1 μM for the tetramer-hexamer equilibrium, and $K_{h\text{-}24} \approx 4000$ μM^3 for the hexamer-24mer equilibrium. Iterative cycles of simulations varying the $K_{m\text{-}d}$ and $K_{d\text{-}t}$ values suggest that these parameters are highly correlated, but independent of $K_{t\text{-}h}$ and $K_{h\text{-}24}$, which are poorly correlated to each other. In other words, a poor agreement between experiment and simulation upon choosing a particular $K_{t\text{-}h}$ value cannot be compensated for by the systematic adjustment of $K_{h\text{-}24}$, nor vice versa.

A comparison of the $K_{m\text{-}d}$ and $K_{d\text{-}t}$ values extracted for ATPγS to those of the ADP-mediated assembly process suggests that the formation of dimers and tetramers may be slightly weaker with Mg·ATPγS than with Mg·ADP (6 and 2 μM vs. 3.5 and 1 μM, see Figure 1). However, the significance of these differences is difficult to evaluate, as a rigorous error estimate is not available. Regardless, only the processes of hexamerization ($K_{t\text{-}h}$) and large aggregate formation ($K_{h\text{-}24}$) are significantly affected by the change in nucleotide (Figure 1). Formation of hexameric species appears substantially promoted by Mg·ATPγS, as the extracted $K_{t\text{-}h}$ value of 0.1 μM is ten-fold smaller than that for Mg·ADP (1 μM). At the same time, assembly to very large aggregates is substantially disfavored in

the presence of Mg·ATPγS, as indicated by a $K_{h\text{-}24}$ value of about 4000 μM^3 compared to only 25 μM^3 for Mg·ADP.

The extracted K_d values were utilized to calculate fractional concentrations of assembly intermediates as a function of concentration (Figure 1, lower panels). Between 1 and 4 μM β-Rca-AC, our model predicts 15 - 18% dimeric species in the presence of Mg·ATPγS, somewhat less than the 26 - 28% predicted for Mg·ADP. However, between 8 and 70 μM, about 60 - 80% hexameric species are predicted for Mg·ATPγS. Under these conditions, the hexameric component reaches a maximum of 76 % at 30 μM β-Rca-AC subunit concentration. These values are significantly higher than the 30-40 % hexamer predicted for Mg·ADP over a similar concentration range. However, at 100 μM β-Rca-AC, our model predicts only 40% high-MW aggregates for Mg·ATPγS, whereas ADP provides for 70% aggregates. In combination, our data clearly indicate that Mg·ATPγS stabilizes the formation of hexameric species, whereas Mg·ADP disfavors hexamers and enhances the formation of larger complexes.

Hydrolysis-deficient β-Rca-AC-D173N binds nucleotides and assembles like wild-type in a nucleotide-dependent fashion. To account for any assembly artifacts due to the substrate analog ATPγS, control experiments were carried out using the natural substrate ATP in combination with the mutant protein β-Rca-AC-D173N. In this variant, Walker-B Asp173 (CLFIN<u>D</u>), predicted to be the catalytic base in ATP hydrolysis by sequence homology (*16*), is replaced by an asparagine residue. Labeling of β-Rca-AC-D173N with AlexaFluor-546 in ATP-containing buffer provided molar ratios of 1.1 ± 0.1 (n = 2) covalently attached fluorophores per chain. Similarly, in ADP-containing buffers, molar

ratios of 0.9 ± 0.1 (n = 3) were determined by use of all analytical methods described above (see also Supplementary Figures S1 and S2). Therefore, this variant was judged to be suitable for FCS studies.

As a first step, ADP-mediated assembly of β-Rca-AC-D173N was compared to that of its parent protein (*34*) (Figure 2). Although D_{app}/D_1 displayed larger standard deviations for the mutant than those typically observed for wild-type, assembly appeared to proceed in nearly identical fashion. Optimization of K_d values using the algorithm described previously provided binding constants within experimental error of those obtained for the wild-type protein (Figure 2, bottom). For β-Rca-AC-D173N with Mg·ADP, the extracted values for $K_{m\text{-}d}$, $K_{d\text{-}t}$, $K_{t\text{-}h}$ and $K_{h\text{-}24}$ were 6 μM, 0.5 μM, 0.5 μM and 5 μM^3 respectively, which compares to 3.5 μM, 1 μM, 1 μM and 25 μM^3 for β-Rca-AC (Figure 2) (*34*). In both cases, the fractional hexamer content appears to reach a maximum at 10 - 15 μM Rca, where it comprises about 40% of all subunits. These results imply that the assembly mechanism is not altered due to the D173N substitution, at least not within the limits of interpretation imposed by the rather noisy FCS data.

Second, we tested whether elevated ATP would favor smaller oligomeric states as observed for the wild-type protein. To this end, β-Rca-AC-D173N samples were prepared in buffers containing three different ATP:ADP ratios at a total nucleotide concentration of 2 mM and a total Mg^{2+} concentration of 5 mM (free Mg^{2+} = 3.0 to 3.1 mM). Surprisingly, ATPase assays carried out in parallel suggested residual ATP hydrolytic activity of the mutant protein. Therefore, kinetic assays were set up on ice to exactly mimic the sample preparation conditions employed for FCS experiments, and used to estimate the average ATP:ADP ratio present in protein samples during FCS data collection (Supplementary

Table S1). In this way, ratios of 3.6 (range 3-6), 2.3 (range 2-3) and 0 (all ADP) were determined to accurately describe the three types of conditions examined (Figure 3). Computer simulations of D_{app}/D_1 versus protein concentration between 50 nM and 8 μM yielded $K_{m\text{-}d}$ values ranging from 6 to 10 μM, and $K_{d\text{-}t}$ values ranging from 0.5 to 2 μM for the three ATP:ADP ratios. For both the monomer-dimer and the dimer-tetramer equilibrium, the all-ADP condition provided the smallest K_d values, again suggesting that ADP may promote a slightly increased propensity towards dimer and tetramer formation (Figure 3). Although this effect is consistent with the data collected on β-Rca-AC (Figures 1 and 2), the differences are rather subtle and difficult to quantitate. To the contrary, extracted $K_{t\text{-}h}$ values for the formation of hexamers exhibited strong ATP-dependence. These values were 0.05 μM, 0.2 μM and 0.5 μM for high, medium and low ATP respectively, revealing increasingly tight hexamerization with increasing ATP:ADP ratios (Figure 3). As observed for the wild-type protein (Figure 1), the estimated binding constant for hexamer formation differed by one order of magnitude between the high and low ATP conditions. The fractional concentrations of oligomeric species calculated from the optimized K_d values suggest that the maximum hexamer content is about 80%, 70% and 60% at the respective ATP:ADP ratios of 3.6, 2.3 and 0. In combination, these data provide strong support for our findings for the assembly process of β-Rca-AC in the presence of ATPγS (Figure 1).

To further examine the effect of nucleotide ratio on assembly, we carried out FCS experiments at constant protein concentration of 60 μM β-Rca-AC-D173N while varying the nucleotide ratio. As before, samples were prepared at 2 mM total nucleotide and 5 mM total Mg^{2+} (free Mg^{2+} = 3.0 - 3.1 mM), but at five different ATP:ADP ratios that consisted

of all-ATP, 3:1, 1:1, 1:3 and all-ADP (Figure 4A). At high ATP, β-Rca-AC-D173N demonstrated a large diffusion coefficient consistent with smaller assemblies of 6-12 subunits. Upon reduction of the ATP content, a slow and gradual decrease in D_{app} is observed, indicating the formation of somewhat larger species, such as a dodecamer at an ATP:ADP ratio of 1/3. However, when ATP is removed completely, a steep drop of D_{app}/D_1 is observed, in line with high-MW aggregate formation significantly larger than a dodecamer. These findings agree with all protein concentration-dependent studies presented above on β-Rca-AC (Figures 1 and 2), confirming that smaller oligomers are favored at higher ATP:ADP ratios.

Free Mg^{2+} promotes hexamer-like oligomeric forms of β-Rca-AC-D173N. To assess whether excess Mg^{2+} affects the Rca assembly process, we used FCS to determine the D_{app} of preparations containing 45 μM β-Rca-AC-D173N in the presence of 2 mM total nucleotide. For this experiment, the ratio of ATP:ADP was kept constant at 3:1, whereas the concentration of total Mg^{2+} was adjusted to provide free Mg^{2+} concentrations (i.e. not coordinated to ATP or ADP) between 0.1 and 5 mM (Figure 4B). Strikingly, the value of D_{app}/D_1 increased substantially with increasing free Mg^{2+}, rising from 0.47 at 0.1 mM free Mg^{2+} to 0.53 at 5 mM free Mg^{2+}. A D_{app}/D_1 value of 0.53 equals the value for a heptameric species as calculated from its MW, and lies within experimental error of the value of 0.55 calculated for a hexameric species (blue bars in Figure 4). Although 6- and 7-fold stoichiometries cannot be distinguished experimentally by FCS, results obtained for the parent protein at the same concentration (45 μM) provide a similar D_{app}/D_1 ratio. For the parent protein, we have interpreted this ratio as representing 70% hexamer and 15% larger aggregates (Figure 1), in support of the promotion of hexamer formation by free

magnesium ions. Notably, the magnesium effect exhibits saturation behavior at concentrations above 3 mM, suggesting specific binding to Rca assemblies in a manner that leads to dissociation of supra-molecular aggregates in favor of hexamers. Therefore, Mg^{2+} is not only a cofactor facilitating ATP binding and hydrolysis, this divalent cation also plays a critical role in mediating the formation of proper subunit contacts, possibly by coordination of acidic groups donated by adjacent chains. We propose that magnesium mediates the assembly of physiologically relevant self-association states.

Thermostability of β-Rca-AC and D173N as a function of nucleotides and Mg^{2+}. We hope that ultimately, a microscopic understanding of Rca assembly states will allow for the design of appropriate studies to characterize the ATPase and Rubisco reactivation activities of functional oligomeric states, as well as examine their 3D structures. In light of the experimental difficulties encountered with the D173N variant, a detailed biophysical characterization of wild-type and substrate-trap variants appears critical not only for the interpretation of FCS data presented here, but also for the design of future experiments. Therefore, the thermostabilities of cotton β-Rca-AC and its D173N variant were examined as a function of substrate analogs and divalent cations. To this end, the Thermofluor assay was employed, an assay that reports on protein unfolding by use of a solvatochromic dye present in solution (*44*). As proteins expose their hydrophobic cores with rising temperature, an increase in fluorescence is induced, allowing for the estimation of the mid-point temperature for protein denaturation ($T_{m\text{-}app}$). Although the extracted $T_{m\text{-}app}$ cannot be considered a thermodynamic quantity, these values allow for the rapid examination of relative stabilities of different expression constructs in different buffer conditions (*24*).

First, the ATP analogues ATPγS and AMP-PNP were tested for their interaction with β-Rca-AC at 2 mM nucleotide (Table 1). Although both analogs were observed to bind Rca, thermal stabilization by ATPγS was substantially more pronounced than stabilization by AMP-PNP, based on the respective $\Delta T_{m\text{-}app}$ values of 44.4 ± 0.2°C and 39.5 ± 0.3°C. ATPγS also provided larger thermal stabilization than ATP ($T_{m\text{-}app}$ = 40.3 ± 0.2°C). These data imply that ATPγS binds tighter than ATP to the Rca active site, and that AMP-PNP destabilizes the protein fold compared to ATP. Surprisingly, the addition of Mg^{2+} to ATPγS samples did not modify the $T_{m\text{-}app}$ values to any extent, and the addition of Mg^{2+} to AMP-PNP samples led to further destabilization by about 1.4°C (Table 1). We conclude that the thermal stabilization previously reported for the addition of Mg^{2+} to ATP-bound Rca is not mimicked by the two ATP analogs ATPγS and AMP-PNP, suggesting electrostatic or structural alterations in the active-site pocket (*24*). Although ATPγS does not appear to modify subunit assembly (Figure 1), AMP-PNP could not be utilized in FCS experiments (see above).

Second, we examined the thermostability of β-Rca-AC-D173N, and compared the values to its parent protein (Table 2). In the absence of nucleotides and Mg^{2+}, the $T_{m\text{-}app}$ value of D173N was shown to be 34.5 ± 0.4°C, somewhat depressed from the β-Rca-AC value of 37.0 ± 0.6°C (*24*), in line with fold destabilization. As with the parent protein, titration of D173N with adenine nucleotide provided an incremental increase in melting temperature, confirming nucleotide binding. Saturation with ADP was shown to raise the $T_{m\text{-}app}$ substantially to 47.5 ± 0.4°C, only slightly below the value of 48.2 ± 0.1°C obtained for the parent (Table 2). The more substantial rise in $T_{m\text{-}app}$ observed for the mutant ($\Delta T_{m\text{-}}$

$_{app}$ = 13°C) compared to the parent ($\Delta T_{m\text{-}app}$ = 11°C) implies stronger ADP binding to D173N. Interestingly, the most stable condition for both β-Rca-AC and β-Rca-AC-D173N appears to be 8 mM ADP in the absence of Mg^{2+} ($T_{m\text{-}app}$ = 48.2°C and 47.5°C, respectively).

When 1 mM free Mg^{2+} was added to 8 mM ADP, the $T_{m\text{-}app}$ decreased for both β-Rca-AC and D173N by 2.7 and 1.2°C respectively (Table 2). However, 1 mM free Mg^{2+} did not affect either of the two apo-proteins, as much as 9 mM free Mg^{2+} was necessary to observe a moderate loss of stability (~1.2 °C). At conditions mimicking the FCS experiments (Figure 2), i.e. 2 mM ADP with 3 mM free Mg^{2+}, the melting temperatures of 42.7°C (β-Rca-AC) and 42.3°C (β-Rca-AC-D173N) indicated essentially identical stabilities for both proteins. Therefore, the larger uncertainties in diffusion coefficients measured for D173N may be a consequence of its propensity to aggregate (Figure 2).

For both proteins, ADP was shown to be more stabilizing than ATP (Table 2). However, this stabilization appeared less pronounced for the mutant ($\Delta T_{m\text{-}app}$ = 4.8°C) than for the parent protein ($\Delta T_{m\text{-}app}$ = 7.6°C) (*24*). At 2 mM ATP and ~1 mM free Mg^{2+} (3 mM total Mg^{2+}), the mutant exhibited a $T_{m\text{-}app}$ value of only 43.8 ± 0.3°C, compared to 45.6 ± 0.2°C for the parent protein. In the absence of Mg^{2+}, the data suggest stronger ATP binding to D173N, likely due to a net reduction in active-site negative charge density that would electrostatically repel the substrate triphosphate tail. In support of electrostatic arguments, addition of Mg^{2+} to ATP stabilizes the wild-type protein ($\Delta T_{m\text{-}app}$ = 4.1°C), but does not provide stabilization of the mutant protein (Table 2).

Residual ATPase activity of β-Rca-AC-D173N, as compared to β-Rca-AC. Although numerous studies on AAA+ proteins have been reported with 'substrate trap' mutants that

lack the catalytic base (*11, 45-50*), a low level of ATPase activity was observed for β-Rca-AC-D173N (Supplementary Table S2). This result was surprising, because previous work on tobacco Rca-D174Q and Rca-D174A did not provide any evidence of ATPase activity (*51*). To be able to interpret FCS data appropriately, we analyzed the ATPase activity of this variant in relation to that of its parent protein by monitoring phosphate production at room temperature via the EnzCheck assay (*39*). At 0.5 and 4 mM ATP, the activity of β-Rca-AC-D173N was observed to be 4- and 13-fold lower than that of the parent protein, whereas bovine serum albumin (BSA) exhibited about 15-fold lower activity (Supplementary Table S2). Since the turnover rates of β-Rca-AC-D173N and BSA were less than one reaction per minute and exhibited a relative standard deviation of 50 – 70%, the residual activity of β-Rca-AC-D173N cannot be distinguished from the non-specific ATPase activity of BSA. These results suggest non-specific, yet protein-mediated substrate hydrolysis (all data corrected for ATP hydrolysis in the absence of protein).

To measure the reduction of the ATP:ADP ratio during sample incubation on ice, substrate hydrolysis was assayed at specific time points over a period of 300 min. For β-Rca-AC-D173N, the EnzCheck assay was utilized to determine the degree of ATP hydrolysis (Supplementary Table S1) (*39*), whereas for β-Rca-AC, an enzyme-linked assay that follows NADH consumption was utilized to determine the degree of ATPγS hydrolysis (*40*). Surprisingly, ATPγS hydrolysis by the parent protein proceeded significantly more slowly than ATP hydrolysis by β-Rca-AC-D173N. For this reason, we were able to collect FCS data on β-Rca-AC (Figure 1) at ATPγS:ADP ratios substantially higher than the ATP:ADP ratios that were achievable for the mutant protein (Figure 3). In separate assays,

we determined that the level of ADP contamination commercially available ATPγS samples is 3.1% -3.4%, whereas the ADP contamination in commercially available ATP samples is only 0.36% - 0.42%. Overall, we conclude that ATPγS-bound β-Rca-AC is more suitable than ATP-bound β-Rca-AC-D173N to address the effects of substrate binding on biophysical behavior, in spite of the relatively large amount of ADP contamination present in ATPγS preparations.

Discussion

Nature often carries out complex biochemical tasks required for cellular homeostasis with the help of macromolecular assemblies that can be extensively regulated (*52, 53*). However, when a protein forms a continuum of self-associated states with subunits that undergo rapid exchange, it becomes increasingly difficult to relate functional data to subunit stoichiometry. Frequently, the exact nature of active oligomeric states and the fractional concentration of these species remain unknown. In AAA+ proteins, hexameric assemblies have been linked to the mechanical remodeling of macromolecular substrates. In addition, experiments on a variety of AAA+ proteins have demonstrated that coupling of ATPase activity with mechanical work is not obligatory (*54*), and can be up- or down-regulated by assembly to higher-order states or by binding to partner proteins (*55*). Both sequential and probabilistic models for subunit conformational change have been proposed, and are thought to involve large-scale structural rearrangements of hexameric rings due to hinge bending motions near the active sites of individual chains. For example, ClpX crystal structures have provided evidence that multiple conformational states can be accessed by individual subunits, with each involving an ensemble of subpopulations (*54*). These features complicate a clear definition of biological activity of each microscopic state.

To begin to address biological functionality in terms of partner protein remodeling, the microscopic equilibria of coexisting quaternary structures must be well-understood.

Cotton short-form Rca forms 70-80% hexamers at four- to eight-fold excess of ATP(γS) over ADP. Building upon previously established FCS methods (*34*), we have characterized the effects of ATP, ATPγS and free Mg^{2+} on the assembly equilibria of cotton short-form β-Rca. Unique to this method is the mathematical framework developed for modeling the thermodynamics of self-association with the explicit inclusion of populated intermediate states. Our work on cotton β-Rca-AC indicates that at low subunit concentrations ranging from 700 nM to 30 μM, the presence of the slowly hydrolyzing ATP analog ATPγS provides larger oligomeric species than the presence of ADP (Figure 1). When utilizing ATPγS, the hexameric form appears to be the dominant species between 4 and 100 μM with a maximum at 30 μM, where ~76% of subunits are predicted to be hexameric. The modeling of states intermediate to the monomeric and hexameric forms was necessary to provide a good fit to the FCS data, suggesting that the dimeric and tetrameric species are well-populated. When the subunit concentration is raised well above 30 μM, ATPγS appears to reduce the contribution of large aggregates in comparison to ADP. This feature indicates that at higher concentrations, ATPγS promotes the formation of hexameric forms as well. For clarity, a summary of all estimated equilibrium constants is presented in Table 3.

Additional experiments carried out with the substrate-trap mutant β-Rca-AC-D173N provided an independent confirmation of the results obtained with wild-type protein. As with β-Rca-AC, D173N preparations with ATP:ADP ratios of 3.6, 2.3 and 0 led to the

prediction of a hexamer peak at ~30 μM subunit concentration. The different nucleotide ratios provided ~ 80%, ~ 70% and ~ 60% hexamer content, respectively (Figure 3), suggesting that ATP:ADP ratios above about 4 may yield maximum hexamer content in *in vitro* assays (Table 3). Unfortunately, β-Rca-AC-D173N appeared somewhat destabilized relative to its parent protein (Table 2), and demonstrated residual hydrolytic activity at a rate somewhat faster than the hydrolysis of ATPγS catalyzed by β-Rca-AC (Tables S1 and S2). Although the affinity for ADP appeared unmodified, D173N appeared to bind ATP more tightly with and without excess magnesium ions, reflecting the reduction in negative charge in the binding pocket (Table 2). The larger error estimates obtained for the mutant D_{app}/D_1 values (Figure 2) indicates an increased tendency towards aggregation, as was also observed during protein purification procedures. Therefore, work with D173N was not pursued further, and averaging of wild-type and mutant K_d values does not appear to be warranted. Therefore, the most reliable dissociation constants may be found in Figure 1 and the wild-type columns in Table 3.

Although many AAA+ proteins such as ClpX and ClpA form tightly bound stable hexameric toroids under most conditions (*13, 47*), a recent study on Vps4 demonstrated that fast-exchanging dimer-hexamer equilibria occur over a broad concentration range of 10 - 200 μM (*48*). Although this observation appears reminiscent of Rca, Vps4 hexamers are more readily formed with ADP than with ATP, an effect opposite to that observed with Rca. The Vpn4 dimer-hexamer K_d value was shown to be 3.7 nM^2 in the presence of ATP and 0.04 nM^2 in the presence of ADP. However, not all Vps4 states observed *in vitro* are thought to be physiologically relevant. For example, a small fraction of larger aggregates appears to form above ~ 60 μM. In addition, the replacement of the Walker B catalytic

Glu has been shown to promote hexameric ring stacking yielding a dodecamer, again an assembly state that is not likely sampled *in vivo* (*48*). Similarly, we propose that the high-MW aggregates observed for Rca·ADP, and to a lesser extent for Rca·ATPγS, may not be biologically relevant states (see below).

The hexamerization K_d values suggest that closed-ring toroidal structures may be promoted at high ATP:ADP ratios at the expense of helical arrangements. In previous work, we proposed a model for β-Rca-AC assembly that consists of three distinct microscopic steps leading to hexameric forms, followed by the formation of larger complexes of ~24 subunits, best modeled as a single step (*34*) (Figure 5). Work presented here demonstrates that the monomer-dimer and dimer-tetramer equilibria exhibit only weak dependence on the ATP:ADP ratio, with $K_{m\text{-}d}$ estimated to be 3 and 6 μM, and $K_{d\text{-}t}$ estimated to be 1 and 2 μM, in the absence and presence of ATPγS respectively (Figure 1 and Table 3). However, substantial differences are observed for hexamer formation, as the $K_{t\text{-}h}$ appears to be 10-fold lower in the presence of ATPγS compared to ADP alone, dropping from 1 μM to 0.1 μM. Since the thermodynamics of hexamerization seems to exhibit a strong dependence on nucleotide, we propose a model in which ATP facilitates the formation of a closed-ring hexameric toroid, whereas ADP facilitates formation of an open hexameric spiral (Figure 5). An optimization of subunit interactions in the presence of ATP is consistent with the expectation that the binding of a dimer to a tetramer would bury twice the exposed molecular surface area, if the ensuing ring forms a closed toroid, rather than an open spiral. We suggest that the lower K_d value reflects the ATP-mediated preference of a doughnut-like arrangement of subunits that would be enthalpically more favorable. On the other hand, in the presence of ADP, the identical $K_{d\text{-}t}$ and $K_{t\text{-}h}$ values of

1 μM may reflect the energetically equivalent addition of a dimer to either another dimer or to a tetramer to form an open spiral arrangement with potential subunit contacts unsatisfied. Therefore, the observed difference in $K_{t\text{-}h}$ values for hexamer formation is consistent with the notion of two competing pathways, such as the parallel formation of both open and closed hexamers. The extracted binding constants likely reflect the relative contributions of each pathway. If this model is correct, the $K_{t\text{-}h}$ values would be predicted to be even lower than 0.1 μM for a system in which only the toroidal hexamer is populated.

Our proposed assembly model (Figure 5) is also consistent with FCS data collected at higher protein concentrations (Figure 1). The $K_{h\text{-}24}$ values describing the assembly of hexamers to supramolecular aggregates were estimated to be 4000 μM^3 in the presence of ATPγS, but only 25 μM^3 in the presence of ADP. This difference in extracted binding constants provides clear evidence that large aggregate formation is substantially diminished in the presence of ATPγS. It is possible that closed-ring hexamers and open hexameric spirals exist in a dynamic equilibrium as a function of the prevailing ATP:ADP ratio (Figure 5). Alternatively, it is also possible that the closed hexamer constitutes an end product, whereas only open hexameric spirals can continue to grow along the helical axis, e.g. by the sequential addition of dimeric units. Third, our data cannot rule out the possibility of the coaxial stacking of closed hexameric rings to generate larger complexes. Either way, FCS experiments cannot be utilized to distinguish between these models for supramolecular aggregate formation.

Several studies on AAA+ proteins have demonstrated that toroidal hexameric assemblies have a tendency to sample asymmetric conformational states (*10, 54*). Therefore, ring closure may be accommodated by individual chains adopting different

structural sub-states that allow for the optimization of intermolecular contacts (*48*). In this way, variations in the type of bound nucleotide may facilitate hinge-bending motions coupled to the reaction cycle, and the thermal accessibility of a large variety of different substates may be necessary for proper function (*48*). A corollary of a biologically driven mandate for cyclic conformational change is that incorrectly assembled states may form more readily in the test tube. This effect may be more pronounced in the presence of non-biological nucleotides, or if the nucleotide is restricted to one type only (see below).

To generate tightly bound hexameric assemblies, subunit contacts must be mediated by magnesium ions. Although Mg^{2+} coordination to the ATP β- and γ-phosphoryl groups has long been known to aid in substrate binding and hydrolysis, the role of Mg^{2+} in promoting productive assembly remains poorly understood. For 45 μM cotton β-Rca-AC-D173N at a ratio of 3:1, ATP:ADP, low Mg^{2+} leads to supra-molecular complex formation, whereas Mg^{2+} in excess of nucleotide coordination favors smaller oligomers consistent with hexameric forms (Figure 3). Strikingly, the estimated dissociation constant for binding of the second magnesium ion lies between 1 and 3 mM, likely within the range of physiological Mg^{2+} fluctuations in the chloroplast stroma believed to occur in response to light/dark adaptation (*23, 56*). Therefore, we propose that *in vivo*, Rca activity may be adjusted by magnesium-mediated changes in subunit contacts. To the best of our knowledge, magnesium ions have not previously been implicated in mediating chain-chain interactions in other AAA+ proteins, although Mg^{2+} activation of Rca hydrolytic activity has been noted some years ago (*25*).

Although a structural model of ATP-bound Rca remains unavailable, the apo-structure of the tobacco Rca AAA+ domain (PDB ID 3T15 and PDB ID 3ZW6) (*30*) reveals the

presence of numerous highly conserved acidic residues near the active site (Figure 6). Some of these, such as Asp288, could participate in the coordination of a second magnesium ion close to a subunit interface, such as by interaction with neighboring-chain Asp242. Although in ClpA, a second Mg^{2+} binding site has been identified on the surface of the N-domain and within 10 Å of bound ADP (*57*), it is not clear whether this site would persist with bound ATP, nor whether the second site has the potential to partake in subunit contacts.

AMP-PNP appears to promote non-native conformations, whereas small amounts of ADP may be critical for functionally relevant ring asymmetry. We have observed that Alexa-dye derivatization in the presence of AMP-PNP results in double-labeling of β-Rca-AC, indicative of partial protein unfolding. In support of this notion, myosin was reported to undergo active site rearrangements in the presence of this analog (*58*). In addition, AMP-PNP appeared to bind only weakly to the cotton β-Rca-AC active site. In the absence of Mg^{2+}, Thermofluor assays suggest that nucleotides interact with Rca with decreasing affinity according to the series ADP > ATPγS > ATP > AMP-PNP (Tables 1 and 2, monitored at 2 mM nucleotide) (*24*). The addition of Mg^{2+} increases the affinity for ATP, decreases the affinity for ADP, has no effect on ATPγS binding and further reduces the affinity for AMP-PNP. Therefore, the presence of excess Mg^{2+} modifies the order to the following series Mg·ATP > Mg·ATPγS > Mg·ADP > Mg·AMP-PNP. We propose that either ADP or AMP-PNP as the sole nucleotide causes a restriction of accessible subunit conformations within the Rca oligomer, in analogy to reports on Vps4 (*48*). This type of mechanism provides a rationale for the preferential formation of non-native states such as spirals, rather than the closed-ring arrangements thought to be biologically relevant. The

ordering of binding strength suggests that active site conformational differences exist that modulate interfacial interactions. In combination, the data presented here support the notion that the simultaneous availability of ATP and ADP may be critical for proper assembly and function. As observed for other ATPases, β-Rca-AC is able to catalyze the slow hydrolysis of ATPγS to ADP (*47, 58*) (Supplementary Table S2), such that a small amount of ADP is produced under all experimental conditions, thereby facilitating biologically relevant ring asymmetry.

Although microscopic binding constants for Rca assembly may vary among different plant species, it is not clear whether these variations are physiologically relevant. Although species-specific differences in Rca assembly remain poorly understood, it appears likely that significant differences in K_d values exist. For example, SAXS data reported for tobacco apo-β-Rca at 2.4 μM were consistent with hexameric assemblies (*32*), whereas cotton β-Rca-AC appears to require concentrations of ~ 30 μM to form substantial amounts of hexamers. Below 500 nM, our data indicate that the β-Rca-AC monomer is the primary species both in the presence of ATPγS and ADP. By contrast, Stotz et al. (*30*) observed hexameric rings in EM images of tobacco Rca-R294V at approximately 500 nM with either Mg·ATP or Mg·ATPγS. Since no EM images were reported for the parent protein, the data suggest that the R294V substitution facilitates hexamerization at very low concentrations. Recently, we have introduced the homologous R293V substitution into the cotton β-Rca-AC construct, and found that this mutation causes partial denaturation and renders the protein incapable of nucleotide binding (Wachter, unpublished data). Therefore, it appears that the R294V effect is relatively specific to tobacco Rca.

The Rca subunit concentration in the chloroplast stroma has been estimated to be 7 to 21 mg/ml (160 to 480 μM) (*22*), whereas that of the Rubisco holoenzyme is thought to be on the order of 500 μM. These values allow for a crude estimate of one Rca hexamer per 10 Rubisco holoenzymes under physiological conditions, where only about 3% of all protein entails Rca, while about 40% consists of Rubisco. Strikingly, stromal Rca concentrations are thought to lie significantly above 100 μM, the highest concentration accessible in our FCS experiments. Since hexamer formation is essentially complete at 30 μM (Figure 1), it appears that *in vivo*, Rca is likely always fully assembled. Therefore, Rca assembly in and of itself may not be directly relevant in the regulation of Rubisco activity. Rather, the adjustment of subunit interactions via nucleotide and magnesium ions may play a critical regulatory role.

Conclusions.

The FCS methodology employed in this study allows, for the first time, for a thermodynamic description of protein self-association in highly polydisperse Rca preparations. From these values (Figure 1), fractional concentrations of oligomeric species can be estimated under a variety of solution conditions. We demonstrate that higher ratios of ATP:ADP facilitate the formation of hexameric species, and show that free Mg^{2+} plays a role in promoting hexamer formation at the expense of larger complexes. We hope that this information will be useful in the interpretation of functional data, and expect that continued application of this technique will allow for a comprehensive analysis of the factors governing Rca assembly, with the ultimate goal of gaining a better understanding of its interaction with Rubisco.

References

1. Spreitzer, R. J., and Salvucci, M. E. (2002) Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol. 53*, 449-475.

2. Parry, M. A., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., and Whitney, S. M. (2013) Rubisco activity and regulation as targets for crop improvement. *J. Exp. Biol. 64*, 717-730.

3. Jordan, D. B., and Ogren, W. L. (1981) Species variation in the specificity of ribulose-biphosphate carboxylase-oxygenase. *Nature 291*, 513-515.

4. Mueller-Cajar, O., Stotz, M., and Bracher, A. (2013) Maintaining photosynthetic CO2 fixation via protein remodelling: the Rubisco activases. *Photosynth. Res. 119*, 191-201.

5. Carmo-Silva, A. E., and Salvucci, M. E. (2013) The Regulatory Properties of Rubisco Activase Differ among Species and Affect Photosynthetic Induction during Light Transitions. *Plant Physiol. 161*, 1645-1655.

6. Snider, J., and Houry, W. A. (2008) AAA+ proteins: diversity in function, similarity in structure. *Biochem. Soc. T. 36*, 72-77.

7. Ogura, T., and Wilkinson, A. J. (2001) AAA(+) superfamily ATPases: common structure-diverse function. *Genes to Cells 6*, 575-597.

8. Ogura, T., Whiteheart, S. W., and Wilkinson, A. J. (2004) Conserved arginine residues implicated in ATP hydrolysis, nucleotide sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J. Struct. Biol. 146*, 106-112.

9. Hartman, J. J., and Vale, R. D. (1999) Microtubule disassembly by ATP-dependent oligomerization of the AAA enzyme katanin. *Science 286*, 782-785.

10. Wang, J., Song, J. J., Seong, I. S., Franklin, M. C., Kamtekar, S., Eom, S. H., and Chung, C. H. (2001) Nucleotide-dependent conformational changes in a protease-associated ATPase HslU. *Structure 9*, 1107-1116.

11. Babst, M., Wendland, B., Estepa, E. J., and Emr, S. D. (1998) The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *Embo J. 17*, 2982-2993.

12. Fodje, M. N., Hansson, A., Hansson, M., Olsen, J. G., Gough, S., Willows, R. D., and Al-Karadaghi, S. (2001) Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase. *J. Mol. Biol. 311*, 111-122.

13. Kress, W., Mutschler, H., and Weber-Ban, E. (2007) Assembly pathway of an AAA+ protein: Tracking ClpA and ClpAP complex formation in real time. *Biochemistry 46*, 6183-6193.

14. Lenzen, C. U., Steinmann, D., Whiteheart, S. W., and Weis, W. I. (1998) Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. *Cell 94*, 525-536.

15. Hanson, P. I., and Whiteheart, S. W. (2005) AAA+ proteins: Have engine, will work. *Nat. Rev. Mol. Cell Bio. 6*, 519-529.

16. Iyer, L. M., Leipe, D. D., Koonin, E. V., and Aravind, L. (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol. 146*, 11-31.

17. Wang, J., Song, J. J., Franklin, M. C., Kamtekar, S., Im, Y. J., Rho, S. H., Seong, I. S., Lee, C. S., Chung, C. H., and Eom, S. H. (2001) Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure 9*, 177-184.

18. Page, A. N., George, N. P., Marceau, A. H., Cox, M. M., and Keck, J. L. (2011) Structure and biochemical activities of Escherichia coli MgsA. *J. Biol. Chem.* 286, 12075 - 12085.

19. Simonetta, K. R., Kazmirski, S. L., Goedken, E. R., Cantor, A. J., Kelch, B. A., McNally, R., Seyedin, S. N., Makino, D. L., O'Donnell, M., and Kuriyan, J. (2009) The Mechanism of ATP-Dependent Primer-Template Recognition by a Clamp Loader Complex. *Cell 137*, 659-671.

20. Lee, S. Y., De La Torre, A., Yan, D., Kustu, S., Nixon, B. T., and Wemmer, D. E. (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes Dev. 17*, 2552-2563.

21. Wang, Z. Y., Ramage, R. T., and Portis, A. R. (1993) Mg(2+) and Atp or Adenosine 5'-[Gamma-Thio]-Triphosphate (Atp-Gamma-S) Enhances Intrinsic Fluorescence and Induces Aggregation Which Increases the Activity of Spinach Rubisco Activase. *Biochim. Biophys. Acta 1202*, 47-55.

22. Portis, A. R. (2003) Rubisco activase - Rubisco's catalytic chaperone. *Photosynth. Res. 75*, 11-27.

23. Barta, C., Dunkle, A. M., Wachter, R. M., and Salvucci, M. E. (2010) Structural changes associated with the acute thermal instability of Rubisco activase. *Arch. Biochem. Biophys. 499*, 17-25.

24. Henderson, J. N., Hazra, S., Dunkle, A. M., Salvucci, M. E., and Wachter, R. M. (2013) Biophysical characterization of higher plant Rubisco activase. *Biochim. Biophys. Acta 1834*, 87-97.

25. Lilley, R. M., and Portis, A. R. (1997) ATP hydrolysis activity and polymerization state of ribulose-1,5-bisphosphate carboxylase oxygenase activase - Do the effects of Mg2+, K+, and activase concentrations indicate a functional similarity to actin? *Plant Physiol. 114*, 605-613.

26. Li, C., Wang, D., and Portis, A. R. J. (2006) Identification of critical arginine residues in the functioning of Rubisco activase. *Arch. Biochem. Biophys. 450*, 176-182.

27. Blayney, M. J., Whitney, S. M., and Beck, J. L. (2011) NanoESI mass spectrometry of Rubisco and Rubisco activase structures and their interactions with nucleotides and sugar phosphates. *J. Am. Soc. Mass Spectrum 22*, 1588-1601.

28. Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011) Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature 479*, 194-199.

29. Henderson, J. N., Kuriata, A. M., Fromme, R., Salvucci, M. E., and Wachter, R. M. (2011) Atomic resolution X-ray structure of the substrate recognition domain of higher plant ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase. *J. Biol. Chem. 286*, 35683-35688.

30. Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011) Structure of green-type Rubisco activase from tobacco. *Nat. Struct. Mol. Biol. 18*, 1366-1370.

31. Wachter, R. M., Salvucci, M. E., Carmo-Silva, A. E., Barta, C., Genkov, T., and Spreitzer, R. J. (2013) Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco-Rubisco activase interaction. *Photosynth. Res. 117*, 557-566.

32. Keown, J. R., Griffin, M. D. W., Mertens, H. D. T., and Pearce, F. G. (2013) Small oligomers of ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase are required for biological activity. *J. Biol. Chem. 288*, 20607-20615.

33. Scales, J. C., Parry, M. A. J., and Salvucci, M. E. (2014) A non-radioactive method for measuring Rubisco activase activity in the presence of variable ATP:ADP ratios, including modifications for measuring the activity and activation state of Rubisco. *Photosynth. Res. 119*, 355-365.

34. Chakraborty, M., Kuriata, A. M., Henderson, J. N., Salvucci, M. E., Wachter, R. M., and Levitus, M. (2012) Protein oligomerization monitored by fluorescence fluctuation spectroscopy: Self-assembly of rubisco activase. *Biophys. J. 103*, 949-958.

35. Thompson, N. L. (1991) Lakowicz, J. R., Ed., pp 337, Plenum Press, New York.

36. Muller, J. D., Chen, Y., and Gratton, E. (2003) Fluorescence correlation spectroscopy. *Method Enzymol. 361*, 69-92.

37. Gendron, P. O., Avaltroni, F., and Wilkinson, K. J. (2008) Diffusion coefficients of several rhodamine derivatives as determined by pulsed field gradient-nuclear magnetic resonance and fluorescence correlation spectroscopy. *J. Fluoresc. 18*, 1093-1101.

38. Muller, C. B., Loman, A., Pacheco, V., Koberling, F., Willbold, D., Richtering, W., and Enderlein, J. (2008) Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *EPL 83*.

39. Webb, M. R. (1992) A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc. Nat. Acad. Sci. USA 89*, 4884-4887.

40. Salvucci, M. E. (1992) Subunit interactions of Rubisco activase - polyethylene glycol promotes self-association, stimulates ATPase and activation activities, and enhances interactions with Rubisco. *Arch. Biochem. Biophys. 298*, 688-696.

41. Dawson, R. M. C. (1986) *Data for Biochemical Research*, 3rd ed., Clarendon Press, Oxford.

42. Salvucci, M. E. (2004) Potential for interactions between the carboxy- and amino-termini of Rubisco activase subunits. *FEBS Lett. 560*, 205-209.

43. Pulido, N. O., Salcedo, G., Perez-Hernandez, G., Jose-Nunez, C., Velazquez-Campoy, A., and Garcia-Hernandez, E. (2010) Energetic effects of magnesium in the recognition of adenosine nucleotides by the F-1-ATPase beta subunit. *Biochemistry 49*, 5258-5268.

44. Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem. 357*, 289-298.

45. Weibezahn, J., Schlieker, C., Bukau, B., and Mogk, A. (2003) Characterization of a trap mutant of the AAA plus chaperone ClpB. *J. Biol. Chem. 278*, 32608-32617.

46. Dalal, S., Rosser, M. F. N., Cyr, D. M., and Hanson, P. I. (2004) Distinct roles for the AAA ATPases NSF and p97 in the secretory pathway. *Mol. Biol. Cell 15*, 637-648.

47. Hersch, G. L., Burton, R. E., Bolon, D. N., Baker, T. A., and Sauer, R. T. (2005) Asymmetric interactions of ATP with the AAA+ ClpX6 unfoldase: Allosteric control of a protein machine. *Cell 121*, 1017-1027.

48. Monroe, N., Han, H., Gonciarz, M. D., Eckert, D. M., Karren, M. A., Whitby, F. G., Sundquist, W. I., and Hill, C. P. (2014) The oligomeric state of the active Vps4 AAA ATPase. *J. Mol. Biol. 426*, 510-525.

49. Buck, M., and Hoover, T. R. (2010) An ATPase R-Finger leaves its print on transcriptional activation. *Structure 18*, 1391-1392.

50. Scott, A., Chung, H. Y., Gonciarz-Swiatek, M., Hill, G. C., Whitby, F. G., Gaspar, J., Holton, J. M., Viswanathan, R., Ghaffarian, S., Hill, C. P., and Sundquist, W. I. (2005) *Embo Journal 24*, 3658-3669.

51. van de Loo, F. J., and Salvucci, M. E. (1998) Involvement of two aspartate residues of Rubisco activase in coordination of the ATP-g-phosphate and subunit cooperativity. *Biochemistry 37*, 4621-4625.

52. Goodsell, D. S., and Olson, A. J. (2000) Structural symmetry and protein function. *Annu. Rev. Bioph. Biom. 29*, 105-153.

53. Ali, M. H., and Imperiali, B. (2005) Protein oligomerization: How and why. *Bioorgan. Med. Chem. 13*, 5013-5020.

54. Stinson, B. M., Nager, A. R., Glynn, S. E., Schmitz, K. R., Baker, T. A., and Sauer, R. T. (2013) Nucleotide binding and conformational switching in the hexameric ring of a AAA+ machine. *Cell 153*, 628-639.

55. Vieux, E. F., Wohlever, M. L., Chen, J. Z., Sauer, R. T., and Baker, T. A. (2013) Distinct quaternary structures of the AAA+ Lon protease control substrate degradation. *Proc. Nat. Acad. Sci. USA 110*, E2002-E2008.

56. Ishijima, S., Uchibori, A., Takagi, H., Maki, R., and Ohnishi, M. (2003) Light-induced increase in free Mg2+ concentration in spinach chloroplasts: Measurement of free Mg2+ by using a fluorescent probe and necessity of stromal alkalinization. *Arch. Biochem. Biophys. 412*, 126-132.

57. Xia, D., Esser, L., Singh, S. K., Guo, F., and Maurizi, M. R. (2004) Crystallographic investigation of peptide binding sites in the N-domain of the ClpA chaperone. *J. Struct. Biol. 146*, 166-179.

58. Gulick, A. M., Bauer, C. B., Thoden, J. B., and Rayment, I. (1997) X-ray structures of the MgADP, MgATP gamma S, and MgAMPPNP complexes of the Dichyostelium discoideum myosin motor domain. *Biochemistry 36*, 11619-11628.

Figures and Tables

Table 4.1. Thermofluor-based stability data for β-Rca-AC as a function of nucleotide analogs and magnesium ion concentration. All data were collected in triplicate (n = 3).

ATPγS (mM)	**AMP-PNP (mM)**	**Mg^{2+} (mM)**	**β-Rca-AC $T_{m\text{-}app}$ (°C ± Std. Dev.)**
			37.0 ± 0.6
		5	35.7 ± 0.4
0.01			37.8 ± 0.3
0.1			39.3 ± 0.3
1			43.1 ± 0.2
2			44.4 ± 0.2
4			45.4 ± 0.1
0.01		5	36.5 ± 0.1
0.1		5	38.9 ± 0.2
1		5	43.4 ± 0.2
2		5	44.5 ± 0.1
4		5	45.7 ± 0.2
	0.01		36.7 ± 0.3
	0.1		36.9 ± 0.2
	1		38.5 ± 0.4
	2		39.5 ± 0.2
	4		40.6 ± 0.3
	0.01	5	35.5 ± 0.1
	0.1	5	35.6 ± 0.1
	1	5	37.1 ± 0.2
	2	5	38.1 ± 0.2
	4	5	39.1 ± 0.3

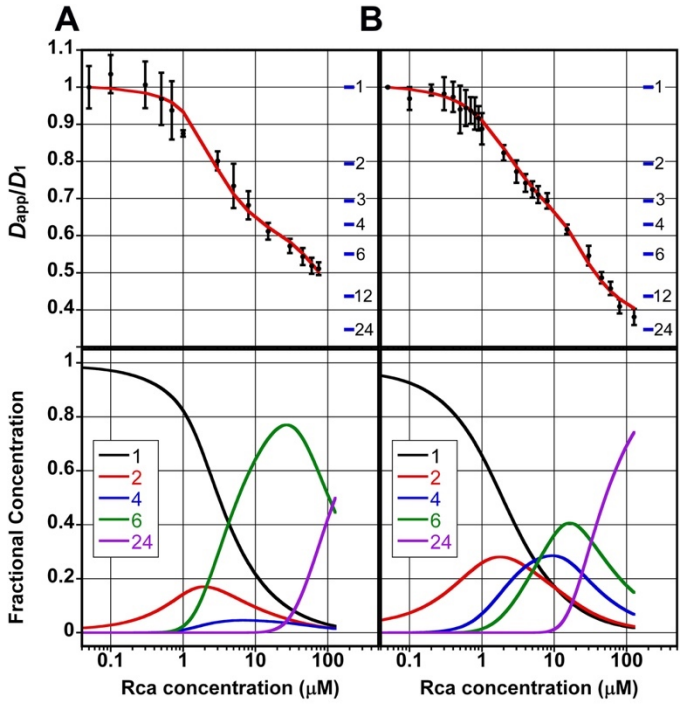
Table 4.2. Thermofluor-based stability data comparing β-Rca-AC and β-Rca-AC-D173N as a function of substrate and product nucleotide, and magnesium ion concentrations. All data were collected in triplicate (n = 3).

ADP (mM)	ATP (mM)	Mg^{2+} (mM)	$T_{m\text{-}app}$ (°C ± Std. Dev.)	
			β-Rca-AC	β-Rca-AC-D173N
0			37.0 ± 0.3	34.5 ± 0.3
0.01			38.4 ± 0.2	36.0 ± 0.1
0.1			41.0 ± 0.2	40.0 ± 0.4
1			45.2 ± 0.1	44.9 ± 0.4
2			46.0 ± 0.2	45.4 ± 0.1
4			47.2 ± 1.1	46.3 ± 0.2
8			48.2 ± 0.1	47.5 ± 0.4
0		0	37.0 ± 0.6	34.5 ± 0.4
1		2	44.5 ± 0.1	44.5 ± 0.1
2		3	42.7 ± 0.2	42.3 ± 0.1
4		5	46.2 ± 0.3	46.6 ± 0.7
8		9	45.5 ± 0.4	46.3 ± 0.1
	0		37.0 ± 0.6	34.5 ± 0.4
	0.01		37.1 ± 0.6	36.9 ± 0.4
	0.1		38.4 ± 0.2	40.9 ± 0.1
	1		40.2 ± 0.2	43.6 ± 0.4
	2		40.2 ± 0.5	44.0 ± 0.2
	4		41.3 ± 0.1	44.2 ± 0.4
	8		40.6 ± 0.1	42.7 ± 0.2
	0	0	37.0 ± 0.6	34.5 ± 0.4
	1	2	45.8 ± 0.1	44.2 ± 0.2
	2	3	45.6 ± 0.2	43.8 ± 0.3
	4	5	45.0 ± 0.5	43.4 ± 0.1
	8	9	44.7 ± 0.2	42.8 ± 0.6
		0	36.7 ± 0.1	34.5 ± 0.4
		1	36.5 ± 0.3	34.8 ± 0.3
		2	35.7 ± 1.1	34.5 ± 0.7
		3	35.7 ± 0.6	33.6 ± 0.4
		5	35.4 ± 0.5	33.6 ± 0.8
		9	35.4 ± 0.1	33.3 ± 0.4

Table 4.3. Summary of estimated equilibrium constants. WT (wild-type) refers to cotton β-Rca-AC, and D173N refers to cotton β-Rca-AC-D173N.

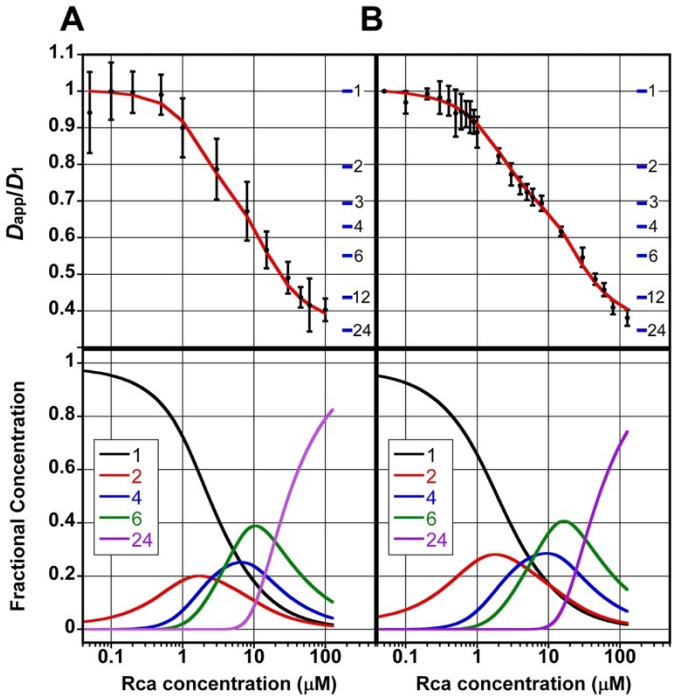
ATP(γS) : ADP	~ 8 : 1		~ 4 : 1		~ 2 : 1		0 : 1	
	WT	D173N	WT	D173N	WT	D173N	WT	D173N
$K_{m\text{-}d}$ (μM)	6	---	---	10	---	10	3.5	6
$K_{d\text{-}t}$ (μM)	2	---	---	2	---	2	1	0.5
$K_{t\text{-}h}$ (μM)	0.1	---	---	0.05	---	0.2	1	0.5
$K_{h\text{-}24}$ (μM^3)	4000	---	---	---	---	---	25	5

Figure 4.1: Assembly mechanism of β-Rca-AC as a function of ATPγS and ADP. *Top Panels:* Relative apparent diffusion coefficients (D_{app}/D_1) of β-Rca-AC determined by FCS as a function of subunit concentration (black circles). Each sample was prepared in buffer containing 25 mM HEPES-NaOH pH 7.6, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 2 mM ATPγS (A) or 2 mM ADP (B). For (A), the ATPγS:ADP ratio was estimated to be 8 (range 7-11) during FCS data collection. The data shown in (B) have been published previously (*34*), and are replotted for ease of comparison. The horizontal blue lines are placed at calculated values of D_{app}/D_1 for k = 1, 2, 3, 4, 6, 12 and 24, where k represents the number of subunits per assembly. The red curves represent computer simulations of the Rca assembly mechanism calculated using the K_d values shown in the table below. *Bottom Panels*: Fractional concentrations of oligomeric species as a function of Rca subunit concentration, as calculated from the K_d values determined by manual curve fitting.



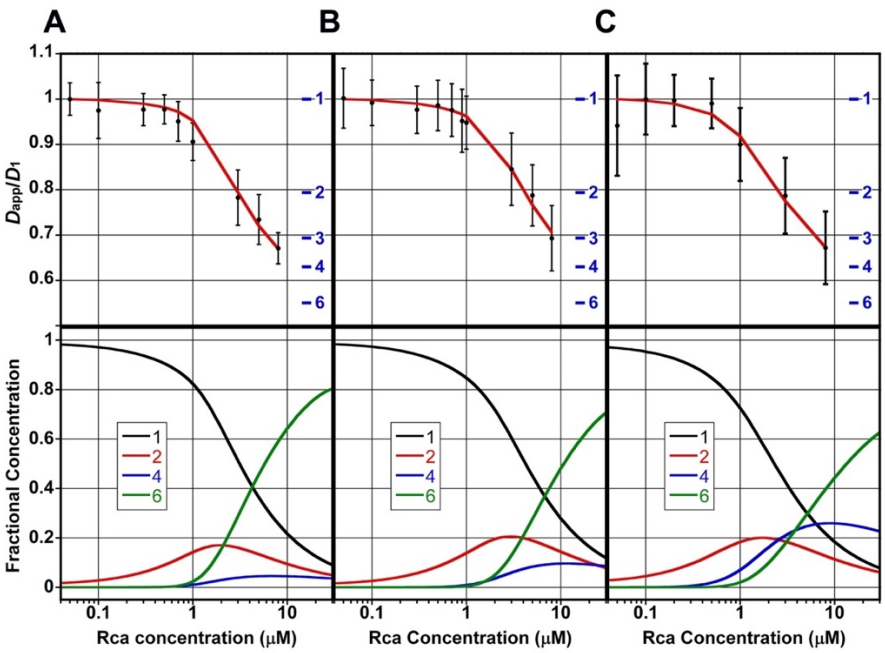
ATPγS / ADP	range 8 (7-11)	0
$K_{m\text{-}d}$ (μM)	6	3.5
$K_{d\text{-}t}$ (μM)	2	1
$K_{t\text{-}h}$ (μM)	0.1	1
$K_{h\text{-}24}$ (μM^3)	4000	25

Figure 4.2: Comparison of the assembly mechanism of β-Rca-AC and β-Rca-AC-D173N protein in the presence of Mg·ADP. *Top Panels:* Relative apparent diffusion coefficients (D_{app}/D_1) determined by FCS as a function of subunit concentration (black circles). Each sample was prepared in buffer containing 25 mM HEPES-NaOH pH 7.6, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 2 mM ADP. (A), β-Rca-AC-D173N, and (B), β-Rca-AC. The data shown in (B) have been published previously (*34*), and are re-plotted for ease of comparison. The horizontal blue lines are placed at calculated values of D_{app}/D_1 for k = 1, 2, 3, 4, 6, 12 and 24, where k represents the number of subunits per assembly. The red curves represent computer simulations of the assembly mechanism calculated using the K_d values shown in the table below. *Bottom Panels*: Fractional concentrations of oligomeric species as a function of Rca subunit concentration, as calculated from the K_d values determined by manual curve fitting.



	D173N + ADP	WT + ADP
$K_{m\text{-}d}$ (μM)	6	3.5
$K_{d\text{-}t}$ (μM)	0.5	1
$K_{t\text{-}h}$ (μM)	0.5	1
$K_{h\text{-}24}$ (μM^3)	5	25

Figure 4.3: Assembly mechanism of β-Rca-AC-D173N as a function of the ATP:ADP ratio. *Top Panels:* Relative apparent diffusion coefficients (D_{app}/D_1) determined by FCS as a function of subunit concentration (black circles). Each sample was prepared in buffer containing 25 mM HEPES pH 7.6, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 2 mM total nucleotide prepared at ratios of (A) ATP:ADP ≈ 3.6, (B) ATP:ADP ≈ 2.3, and (C) all ADP. The horizontal blue lines are placed at calculated values of D_{app}/D_1 for k = 1, 2, 3, 4, 6, 12 and 24, where k represents the number of subunits per assembly. The red curves represent computer simulations of the Rca assembly mechanism calculated using the K_d values shown in the table below. *Bottom Panels*: Fractional concentrations of oligomeric species as a function of β-Rca-AC-D173N subunit concentration, as calculated from the K_d values determined by manual curve fitting. Note, for panels A and B, estimated average ATP:ADP ratios of 3.6 (range 3 – 6) and 2.3 (range 2 – 3) are reported based on the measured kinetics of ATP hydrolysis during sample preparation.



	A	B	C
ATP/ADP (range)	3.6 (3-6)	2.3 (2-3)	0
$K_{m\text{-}d}$ (μM)	10	10	6
$K_{d\text{-}t}$ (μM)	2	2	0.5
$K_{t\text{-}h}$ (μM)	0.05	0.2	0.5

Figure 4.4: (A) Formation of supramolecular assemblies of 60 μM β-Rca-AC-D173N with decreasing ATP:ADP ratios. Apparent (D_{app}) and normalized diffusion coefficients (D_{app}/D_1) were determined in the presence of different nucleotide ratios as indicated, while keeping the protein concentration constant. The error bars represent the standard deviations of at least three independent measurements. (B) Disassembly of 45 μM β-Rca-AC-D173N with increasing free Mg^{2+} concentrations. Apparent (D_{app}) and normalized diffusion coefficients (D_{app}/D_1) were determined in presence of 1.5 mM ATP and 0.5 mM ADP (ATP:ADP=3:1). The total Mg^{2+} concentration was utilized to calculate free Mg^{2+}, i.e. Mg^{2+} not coordinated to nucleotide. The error bars represent standard deviations of experiments carried out in triplicate.

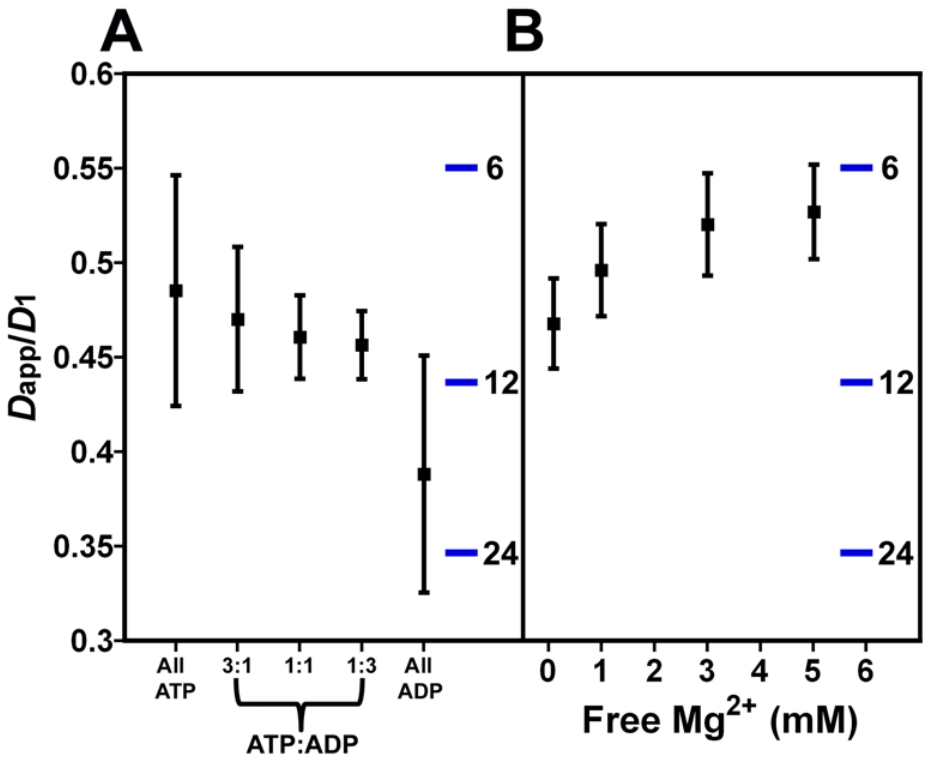


Figure 4.5: Schematic representation of the proposed Rca assembly model. Individual chains of the Rca AAA+ domain are shown in light and dark blue. $K_{m\text{-}d}$, monomer-dimer dissociation constant (K_d), $K_{d\text{-}t}$, dimer-tetramer K_d, $K_{t\text{-}h}$, tetramer-hexamer K_d, and $K_{h\text{-}24}$, hexamer-24mer K_d value. According to this model, dimeric and tetrameric oligomers assemble to form hexameric species. Two different kinds of assemblies consisting of six subunits each are proposed to be generated *in vitro*, a closed-ring and open-spiral form. The relative populations of each hexamer type are proposed to depend on nucleotide and magnesium conditions, and the experimentally determined $K_{t\text{-}h}$ value is thought to reflect the relative contributions of each species (lower $K_{t\text{-}h}$ indicates more pronounced closed-ring hexamer formation). The figure was generated by calculation of a coarse surface of the crystal structure of the AAA+ domain of tobacco Rca (pdb ID code 3T15), using the program Pymol.

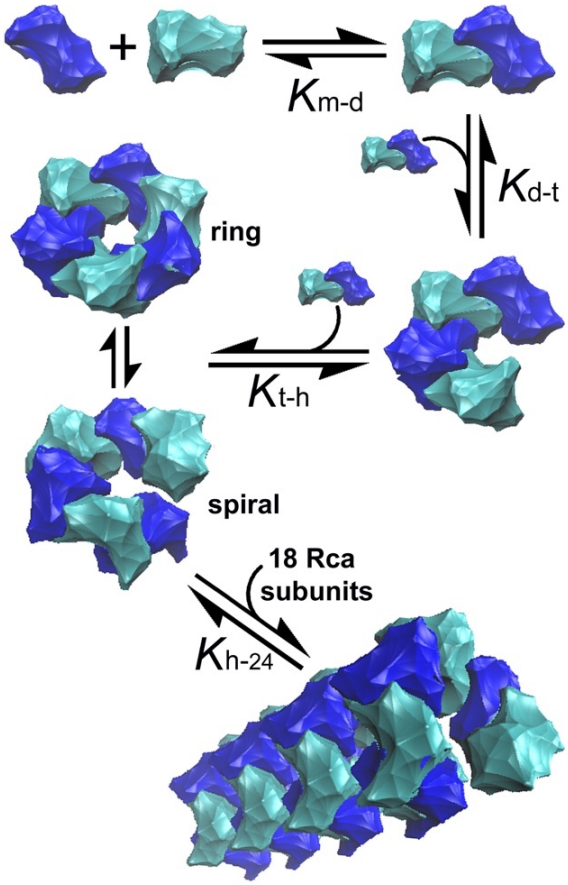
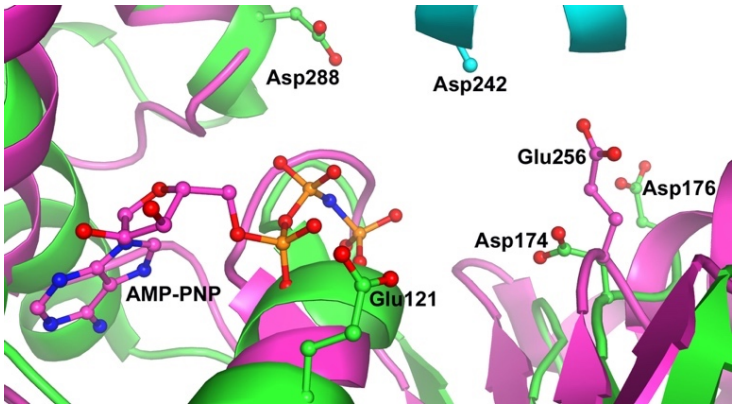


Figure 4.6: Structural superimpositions of nucleotide-bound FtsH and tobacco Nt-apo-β-Rca. *Magenta:* Cartoon depicting the structure of the AMP-PNP-bound AAA+ module from FtsH (PDB ID 1IY0). The ATP analog AMP-PNP and the catalytic Glu256 of the Walker B motif are represented in ball-and-stick format. *Green:* AAA+ module of Nt-apo-β–Rca (PDB ID 3T15) with the catalytic Asp174 of the Walker B motif and additional conserved acidic residues (Glu121, Asp176 and Asp288) represented in ball-and-stick format. *Cyan:* Adjacent Nt-apo-β-Rca protomer at the crystallographic interface showing the proximity of conserved Asp242 (side chain modeled to Cβ only) to the nucleotide binding pocket. Figure prepared with the PyMOL Molecular Graphics System, Schrödinger, LLC.



CHAPTER 5

SINGLE-MOLECULE DIFFUSOMETRY REVEALS THE NUCLEOTIDE-DEPENDENT OLIGOMERIZATION PATHWAYS OF NICOTIANA TABACUM RUBISCO ACTIVASE

Work presented in this chapter reflects previously published work by Wang *et al.* As a co-author on this publication, I am including it as part of my thesis, given that I have made significant contributions to this work. This work was part of a collaboration between the Wachter group at Arizona State University and the Moerner group at Stanford University. The nature of the collaboration was such that I, as part of the Wachter group would generate the Tobacco Rca sample for Dr. Quan Wang, a post-doc in the Moerner lab at the time, to use on their ABEL Trap set-up. To that end, I purified recombinantly expressed tobacco Rca and labeled it with the Alexa 647 dye required by the ABEL Trap. I performed all the necessary characterization of the labeled sample to ensure that it was of high enough quality for their system. Characterization of the sample included RP-HPLC and MALDI-TOF work. I also purified and labeled a true wild-type sample which lacked the C-terminal tail mutation to determine whether our label was indeed attaching to the desired cysteine. In addition, I designed the system for aliquoting the sample based on the research questions we wanted to answer and prepared all necessary buffers for the experiments. In terms of the writing of the manuscript, I contributed by writing the methods section regarding the protein purification and characterization, edited the final manuscript prior to submitting, and provided feedback when the reviewers of the manuscript made comments.

The full reference for the article is given below:

Wang, Q., *et al*. 2018. Single-molecule diffusometry reveals the nucleotide-dependent oligomerization pathways of *Nicotiana tabacum* Rubisco activase. *J. Chem. Phys.* **148,** 123319. DOI: 10.1063/1.5005930

ABSTRACT

Oligomerization plays an important role in the function of many proteins, but a quantitative picture of the oligomer distribution has been difficult to obtain using existing techniques. Here we describe a method that combines sub-stoichiometric labeling and recently-developed single-molecule diffusometry to measure the size distribution of oligomers under equilibrium conditions in solution, one molecule at a time. We use this technique to characterize the oligomerization behavior of *Nicotiana tabacum* (Nt) rubisco activase (Nt-Rca), a chaperone-like, AAA-plus ATPase essential in regulating carbon fixation during photosynthesis. We directly observed monomers, dimers and a tetramer/hexamer mixture, and extracted their fractional abundance as a function of protein concentration. We show that the oligomerization pathway of Nt-Rca is nucleotide dependent: ATPγS binding strongly promotes tetramer/hexamer formation from dimers and results in a preferred tetramer/hexamer population for concentrations in the 1-10μM range. Furthermore, we directly observed dynamic assembly and disassembly processes of single complexes in real time, and from there estimated the rate of subunit exchange to be ~$0.1s^{-1}$ with ATPγS. On the other hand, ADP binding destabilizes Rca complexes by enhancing the rate of subunit exchange by >2 fold. These observations provide a quantitative starting point to elucidate the structure-function relations of Nt-Rca complexes. We envision the method to fill a critical gap in defining and quantifying protein assembly pathways in the small-oligomer regime.

Introduction

The oligomerization state of many proteins is critical to their function[1]. It forms the basis for substrate recognition for many enzymes[2] and underpins the rich biophysical phenomena of allostery[3] and cooperativity[4,5]. On the other hand, erroneous aggregation is involved in the pathogenesis of various amyloid diseases (e.g. Alzheimer's, Parkinson's, Huntington's, etc.)[6,7]. Understanding the mechanism and pathways by which proteins assemble into higher order structures is of fundamental and biomedical significance. Several biochemical and biophysical methods are available to examine the oligomerization behavior of proteins but all have certain limitations. Methods based on liquid chromatography[8] and sedimentation[9,10] result in at least partial fractionation of the sample, an effect that is not suitable for complexes that undergo rapid exchange during measurements. Scattering-based techniques like small angle X-ray scattering (SAXS)[11] and dynamic light scattering (DLS)[12] rely on model-dependent fitting, which can be error-prone for highly polydispersed and/or interacting samples. Recently, native mass spectrometry has shown great promise in characterizing the quaternary structure of proteins, but the degree of perturbation introduced by electrospraying macromolecules into the gas phase is still debated[13].

On the other hand, useful methods with single-molecule sensitivity have been developed. Electron microscopy offers structural information at the single-particle level but requires samples to be moved out of equilibrium onto a surface and dried in a vacuum. Recently, fluorescence correlation spectroscopy has been applied to resolve oligomerization states of

proteins[14,15] in solution, but this method has limited resolution for a mixture of assembly states[16]. Other techniques, such as two-color coincidence detection[17], stepwise photobleaching[18] and single-molecule FRET[19] have been applied to characterize oligomerization under the assumption that the perceived fluorescence brightness signal reflects differences in oligomer size. Here, we describe a more direct method to measure the oligomer distribution of proteins in solution, with single-molecule resolution under equilibrium conditions. The method, termed "ABEL-oligo", is based on measuring the hydrodynamic radius of individual molecular oligomers contained in an aqueous-phase Anti-Brownian ELectrokinetic (ABEL) trap in real-time. We use this method to dissect the oligomerization pathway of tobacco Rubisco activase, a catalytic chaperone essential for carbon fixation.

Rubisco activase (Rca) is a chaperone-like, AAA+ ATPase that rescues the main carbon fixation enzyme Rubisco [ribulose-1,5-bisphosphate (RuBP) carboxylase/oxygenase] from premature inhibition[20–22]. Despite its importance, the detailed mechanism of Rubisco activation by Rca is poorly understood, in part due to the complicated oligomerization behavior of Rca observed *in vitro*. Previous biophysical characterization of several higher plant Rca proteins *in vitro* using a suite of established methods[8] revealed a highly heterogeneous and transient oligomer distribution, but these approaches could not be used to establish a quantitative oligomerization pathway. Surprisingly, sedimentation velocity experiments on tobacco (*Nicotiniana tabacum*, Nt) Rca[9,10] suggested that the oligomerization of Nt-Rca does not depend on nucleotide. On the contrary, the assembly of spinach Rca was shown to be strongly dependent on the type of nucleotide[20], consistent

with earlier work reporting that both ATP and ATPγS promote the formation of higher-order oligomers (MW > 550kDa), whereas ADP does not[23]. Additionally, the oligomerization behavior of cotton □-Rca, as monitored by FCS, appeared to be a function of nucleotide.[24]. Meanwhile, several structural studies suggested hexamers as the functional species[22,25,26] (Figure 1B). Here we set out to apply our ABEL-oligo method to define and characterize the assembly pathway of Nt-Rca *in vitro* and examine the effects of ATP binding (with Mg·ATPγS) and hydrolysis (with Mg·ADP).

Materials and Methods

Site-Directed Mutagenesis, Protein Expression and Purification. The gene coding for *Nicotiana tabacum* (tobacco, Nt) Rca (383 amino acid residues, MW 42.764 kDa, theoretical PI 5.5) was transferred into the pHUE plasmid[27,28] following published procedures[29]. The QuikChange mutagenesis kit (Agilent Genomics) was used to introduce the S379C substitution near the C-terminus of Nt-Rca, as confirmed by DNA sequencing. Nt-Rca-S379C was expressed in *E. coli* BL21*(DE3) (Invitrogen) as a 6His-tagged ubiquitin fusion protein, and purified via affinity chromatography on a Ni^{2+}-nitrilotriacetic acid (Ni-NTA) column (Qiagen). Subsequently, the N-terminal 6His-ubiquitin tag was cleaved using a deubiquitylating enzyme, the sample was dialyzed, and reapplied to a Ni-NTA column, as described previously[29]. Protein fractions were spin-concentrated and buffer-exchanged into 25 mM HEPES pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 2 mM ADP. Protein preparations were quantitated using the Bradford method with

bovine serum albumin as a standard. Aliquots were flash-frozen in liquid N_2 and stored at -80°C.

Protein labeling method. The conjugation of the AlexaFluor-647 C_2-maleimide fluorophore (Molecular Probes, Eugene, OR) to the engineered cysteine residue of Nt-Rca-S379C was carried out as described previously[24,29]. Briefly, labeling was carried out in the presence of 5 mM ADP at a dye:protein ratio of 3.5 in HEPES buffer pH 7.2. In control reactions containing wild-type Nt-Rca instead of Nt-Rca-S379C, the covalent labeling of protein was not observed (Figure S1), indicating that the cysteine residue in position 379 is specifically targeted by the maleimide functional group. Alexa-labeled Nt-β-Rca-S379C was purified by gel filtration, then analyzed by reverse-phase HPLC (Agilent Technologies 1100) on a C18 analytical column (Vydac 218 TP) using a linear water/acetonitrile gradient with 0.1% trifluoroacetic acid (TFA). Elution of protein and dye were monitored by optical density (OD) at 220, 280, and 650 nm. Protein eluted at 46.9 % acetonitrile as a single peak and carried the Alexa-dye label, whereas free Alexa-dye was not detected (expected elution at 16.4 % acetonitrile). Alexa/protein ratios were determined by absorbance measurements on collected HPLC fractions as described previously, and provided a 1:1 molar ratio of protein to dye[14]. Aliquots of labeled (88.2 μM) and label-free (268.0 μM) Nt-Rca-S379C in buffer containing 25 mM HEPES−NaOH, pH 7.5, 250 mM KCl, 2 mM ADP, 5 mM $MgCl_2$, 10% glycerol were flash frozen in liquid N_2 and stored at -80˚C.

Sample preparation for ABEL-oligo measurements. On the day of the experiment, samples were thawed on ice. Labeled Rca was first diluted to 3nM concentration in a

buffer containing no Mg^{2+} or nucleotide (50mM HEPES, pH 7.5, 150mM KCl) and kept on ice for ~20min. Subsequently, a reaction mixture containing ~10pM labeled Rca and various amounts (specified in the Results section) of unlabeled Rca was created in the reaction buffer containing Mg^{2+} and nucleotide (50mM HEPES, 150mM KCl, 3mM Trolox, 0 or 5mM $MgCl_2$, 0 or 2mM nucleotide, 10% glycerol). The mixture was incubated on ice for ~20min, deoxygenated (see below) and injected into the ABEL trap for single-molecule measurements. For cotton Rca, the 20min incubation time has been shown to be sufficient to reach equilibrium[14]. The concentration of labeled Rca was kept at ~10pM in all measurements. In experiments with 2mM ATPγS, the final solution also contained between 0 and 80μM ADP, which was carried over from the storage buffer of the unlabeled Rca preparation. Likewise, the low-ADP condition contained between 0 and 80μM ADP.

ABEL trap setup. The ABEL trap was implemented as previously described[30]. Briefly, the apparatus consists of a position-sensing module, a signal-processing unit and a microfluidic sample holder. To sense the position of a single labeled biomolecule or oligomer in real time, we used fast laser scanning in a deterministic pattern and photon-by-photon molecule localization defined by the position of the laser spot at the moment of detection of a fluorescence photon. The signal-processing unit computes the most-probable position of the diffusing molecule given all available physical information. It is implemented on a field programmable gate array (FPGA) board for minimum computational delay. Given estimated position values, feedback voltages were generated, amplified and applied to the solution in a microfluidic sample holder to compensate for the Brownian motion of a single molecule. Our ABEL trap suppresses the Brownian motion

of a single molecule in 2D (xy) only; diffusion in z is confined by the axial thickness of the microfluidic cell to ~700nm.

Single-molecule diffusometry The diffusion coefficient and electrokinetic mobility of individual trapped molecules were extracted using a maximum-likelihood estimator as previously described[31]. This works even though the molecule is prevented from diffusing away, because at any given time, the molecule is still driven by thermal diffusion as well as by the fields used for trapping. Briefly, the algorithm uses the photon-stamped position sequence of a single trapped molecule and the feedback voltages as inputs and utilizes an iterative procedure to reconstruct the motion trajectory. It then separates the in-trap motions of a single molecule into a voltage- dependent component and a stochastic component to estimate electrokinetic mobility and diffusion coefficient, respectively. The algorithm can be configured to run in real-time, or with higher precision offline, with a tunable time resolution between 50ms to ~1s. The estimated diffusion coefficient has a Gaussian distributed uncertainty that scales with (N_{ph} being the number of photons detected during the estimation window, Supplementary Information). In general, both single-molecule hydrodynamic radius (from diffusion coefficient) and charge (from electrokinetic mobility) information can be extracted, as previously demonstrated for DNA hybridization[31]. In this work, the electrokinetic mobility is dominated by the electroosmotic response of the bulk solution (resulting from the polyelectrolyte multi-layer coating (see below) to prevent protein sticking) and contains little information about protein charge. We thus utilized only the diffusion information to identify the different oligomers.

Conditions for single-molecule fluorescence detection. We used the 633nm line from a HeNe laser as the excitation source with an average excitation intensity of 1.5 kW/cm^2 Trapping experiments were performed in the reaction buffer (50mM HEPES, pH 7.5, 150mM KCl, 3mM Trolox, 0 or 5mM $MgCl_2$, 0 or 2mM nucleotide, 10% glycerol) with an added oxygen scavenging system[32] (2mM protocatechuic acid and 50nM protocatechuate 3,4-dioxygenase) to prevent premature photobleaching of the dye Alexa647. For experiments that aimed to probe the dynamics assembly/disassembly processes in real time, the excitation intensity was lowered by a factor of two (0.75 kW/cm^2) and the oxygen scavenging system were made fresh before the measurements. To prevent protein sticking to the microfluidic surfaces, the sample chamber was passivated by polyelectrolyte multilayers using layer-by-layer deposition[33,34]. Sticking is easily detected by observation of the appearance of non-random behavior in the trapping field direction (i.e. the mean value of the feedback voltages being non-zero).

Data Analysis. In the experiment, we only probe protein complexes containing at least one fluorescent subunit. We need to establish a link between the measurable quantities and the underlying oligomer distribution. The probability of an n-mer incorporating exactly one labeled subunit (L=1) assuming independence is given by the binomial distribution,

$$\frac{C_n^{L=1}}{C_n^{tot}} = f\left(1; n, p\right) = \binom{n}{1} p\left(1-p\right)^{n-1} \approx np \qquad (1)$$

where the denominator is the total concentration of an n-mer, and p is the ratio between labeled and total Rca subunits in solution. Labeled Rca was kept at ~10pM and the total unlabeled Rca subunit concentration was varied between 200nM and 11μM. Under these

conditions, p lies between 10^{-6} and 10^{-4} (e.g. 10pM / 200nM = 0.5×10^{-4}), so $(1-p)^{n-1} \approx 1$ and oligomers with more than one labeled subunit can be ignored.

We can now relate the fractional concentration (P^n) of the different oligomers to the apparent percentage concentrations measured using the fluorescent subset.

$$P^n = \frac{nC_n^{tot}}{\sum_k kC_k^{tot}} = \frac{C_n^{L=1}/p}{\sum_k C_k^{L=1}/p} = \frac{C_n^{L=1}}{\sum_k C_k^{L=1}} \tag{2}$$

From the above equation, it becomes clear that by measuring the fluorescent subset of particles and the oligomeric stoichiometries n, we can extract the fractional concentrations of the different oligomers.

The changes in diffusion coefficient were detected by a change-point algorithm similar to Ref.[35] Given that the estimation uncertainty follows a Gaussian distribution[31], we modified the algorithm to use a Gaussian-based log-likelihood ratio test instead of a Poisson model as in the original publication. Two-dimensional densities were estimated from a scatter plot using a Gaussian-smoothing algorithm (2D kernel density estimation[36]).

For a monomer-dimer-tetramer-hexamer assembly model, we have

$$\begin{aligned} K_d^{1-2} &= C_1^{\,2}/C_2 \\ K_d^{2-4} &= C_2^{\,2}/C_4 \\ K_d^{4-6} &= C_2 \cdot C_4 / C_6 \end{aligned} \tag{3}$$

The K_d values were determined by a least-square procedure minimizing the following error function

$$err = \sum_i \left(\tilde{P}_i^1 - P_i^1\right)^2 + \left(\tilde{P}_i^2 - P_i^2\right)^2 + \left[\tilde{P}_i^{>2} - \left(P_i^4 + P_i^6\right)\right]^2 \tag{4}$$

where $\tilde{P}_i^n$ is the experimentally extracted fractional concentration of a n-mer at the i-th concentration point (Figure S3 and Tables S1-S3) and P_i^k is the fractional concentration

predicted by the equilibrium (Eq. 3) at the corresponding protomer concentration, $\tilde{P}_i^{>2}$ represents the unresolved tetramer/hexamer mixture seen in the experiment. This procedure can reliably extract K_d^{1-2} and K_d^{2-4} but not K_d^{4-6} (the error function is insensitive to the value of K_d^{4-6}, Figure S2). As a result, only K_d^{1-2} and K_d^{2-4} are interpreted in this work.

Results

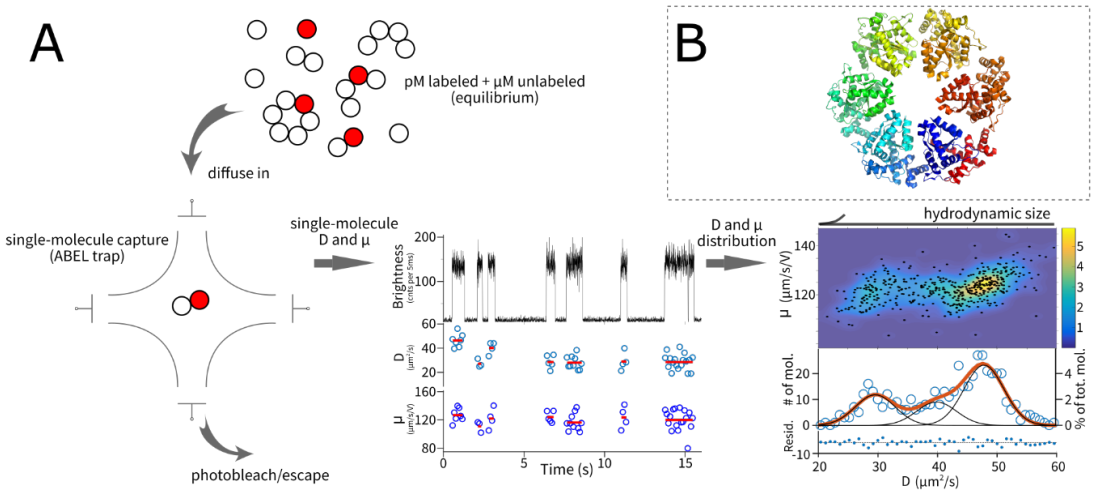


Figure 5.1 Measurement scheme. (A) Workflow of the assay based on single-molecule diffusometry. See main text for details. (B) Structural model of the closed-ring hexamer of truncated Nt-Rca (PDB: 3ZW6). Full-length protein chains were measured in this work.

Resolving oligomers under equilibrium conditions using single-molecule diffusometry

To resolve the different oligomerization states of Nt-Rca in solution, we first incubated labeled Nt-Rca (~10pM) with unlabeled proteins (0-11µM of protomer concentration). After equilibrium was reached, labeled Nt-Rca was randomly incorporated into the different assembly states. We then injected the equilibrated sample into a microfluidic-based single-molecule trapping system known as the Anti-Brownian ELectrokinetic

(ABEL) trap described in Materials and Methods (Figure 1A). Inside the trap, individual fluorescently-tagged proteins were detected in solution with their 2D diffusive motion suppressed to within 2 μm from the trap center, resulting in long-time monitoring of single-molecule behavior in aqueous buffer solution. In the trap, individual proteins enter the center ("trapping region") of the microfluidic chamber by diffusion and, if they carry a fluorescent subunit, are detected and captured by the feedback electrokinetic forces. These trapping events manifested themselves as brightness plateaus lasting ~1 second (Figure 1A). Termination of a trapping event is most likely due to photobleaching of the fluorescent label (Alexa647). Almost all observed molecules (>99%) entered the trap with a brightness that corresponds to one Alexa647 molecule and left the trap with a single photo-bleaching step, supporting the assumption that there is at most one labeled subunit per complex (Materials and Methods).When a single protein was captured, its diffusion coefficient and electrokinetic mobility were estimated with ~100ms time resolution using a parameter estimation framework developed in our previous work[31]. Briefly, the raw particle trajectory in the trap is fitted with a Langevin-type dynamic model with proper treatment of measurement noise.The diffusion coefficient (D) is related to the hydrodynamic radius (r) of the molecule through the Stokes-Einstein relation ($D=k_BT/6\pi\eta r$), and directly probes oligomerization state. Here, the electrokinetic mobility (μ) is only weakly dependent on molecular charge, because trapping was dominated by electro-osmotic forces resulting from charged, anti-fouling surface coatings of the microfluidic chamber (Materials and Methods). As a result, mobility values were expected to be similar for all measured molecules, regardless of their physical properties, as observed in Figures 1A and 2.

The ability to measure single-molecule size allowed us to probe the full distribution of oligomerization states in solution. We thus averaged all *D* values and all μ values collected for each individual protein complex. For all measured molecules, the average values were displayed as single points on a *D*-μ parameter space and also shown as a density distribution plot using a 2D adaptive kernel density estimation method[36]. In the example data shown in Figure 1A, obtained with 400nM Nt-Rca in the presence of ATPγS and Mg^{2+}, we found three components in equilibrium with each other, as the projected *D* histogram require a minimum of three Gaussian distributions to fit. Sub-stoichiometric labeling and single-molecule diffusometry thus provide a powerful means to dissect the distribution of oligomerization states in solution under equilibrium conditions.

The dimer is an Nt-Rca assembly intermediate

We conducted single-molecule diffusometry measurements with varying protein (subunit) concentrations [Rca] from 0-11μM, under three different nucleotide conditions. The results are summarized in Figure 2. Below 200nM, we saw exclusively a single population with a diffusion coefficient of ~50 $\mu m^2/s$, which we identified to be the monomeric species. The value of the assigned monomer species (50 $\mu m^2/s$) is consistent with theoretical modeling of hydrodynamic properties[37] based on the crystal structure (Supplementary Information). This behavior of Nt Rca is similar to cotton β-Rca, which was observed to be exclusively monomeric below 300nM[14].. At intermediate protein concentrations between 0.4μM and 2μM with no Mg^{2+} and low ADP (Figure S3 left), we observed multi-modal distributions, indicating assembly into well-defined oligomeric states. When fitting the projected D histogram (Figure 1A and Figure S3) using three Gaussians, we consistently resolved a

distinct population with a D that is 80% of the monomer. We identified this population to be Rca dimers: to a first-order approximation, the diffusion coefficient of a protein scales with $(MW)^{-1/3}$ (Ref.[38]), where MW is the molecular weight (i.e. for dimers, the diffusion coefficient should be $1/2^{1/3}$=0.79 times that of the monomer, consistent with the 0.8 value observed here). We thus identified the dimer as an assembly intermediate under the low ADP condition. We note that although this simple scaling law between diffusivity and molecular weight is derived by approximating proteins as spheres, a consideration of more realistic protein shapes or structural models has been demonstrated to result in only minor corrections (D_2/D_1 from 0.79-0.80)[14]. Similarly, following the same analysis, dimers were also identified as assembly intermediates under Mg·ATPγS (Figure 3S middle) and Mg·ADP (Figure 3S right) conditions.

Oligomerization of Nt-Rca is nucleotide-dependent

Upon visual inspection, the oligomerization pathways are distinct for the different nucleotide conditions (Figure 2). Here, the diffusion coefficient is normalized by that of the monomer extracted from the fit (Figure 3S, in cases where few monomers were observed, the D of the monomer species was estimated from the dimer population by $D_1=D_2(1/2)^{-1/3}$). To understand these differences more quantitatively, we highlight the expected D positions of the monomer, dimer, tetramer and hexamer species on the D-μ map using the $D_n/D_1 = (1/n)^{1/3}$ scaling law. Again, although these estimates were made by the simple "ball-filling" model, considering realist protein shapes results in very small corrections (<5%, Supplementary Information) In the first condition (Figure 2 left), Mg^{2+} was omitted and residual ADP concentrations ranged from 0 to 80 □M. Since the K_i value

for ADP inhibition of Nt Rca has recently been determined to be 37 □M[39], we estimate that on average, about half of the active sites were occupied with nucleotide. Under these conditions, we observed the buildup of dimers and the appearance of tetramers in the range of 0.4 to 0.8μM Rca. At 2μM, Nt-Rca seemed to exist in a complex equilibrium containing monomers, dimers, tetramers and hexamers. Above 4μM, we observed near complete depletion of the monomer species while most of the population was found to exist as a mixture of tetramers and hexamers. At 11μM, oligomers larger than hexamers were observed. On the other hand, in the presence of 5 mM Mg^{2+} and 2 mM ATPγS, we observed a dramatically different oligomerization progression. Between 0.4 and 0.8μM, we observed weak dimer buildup and rapid formation of a D~0.6D_1 species. We identified this state to be a mixture of tetramers and hexamers, given the limited resolution in measuring D values[31] compared to the relatively small difference in D between tetramers and hexamers

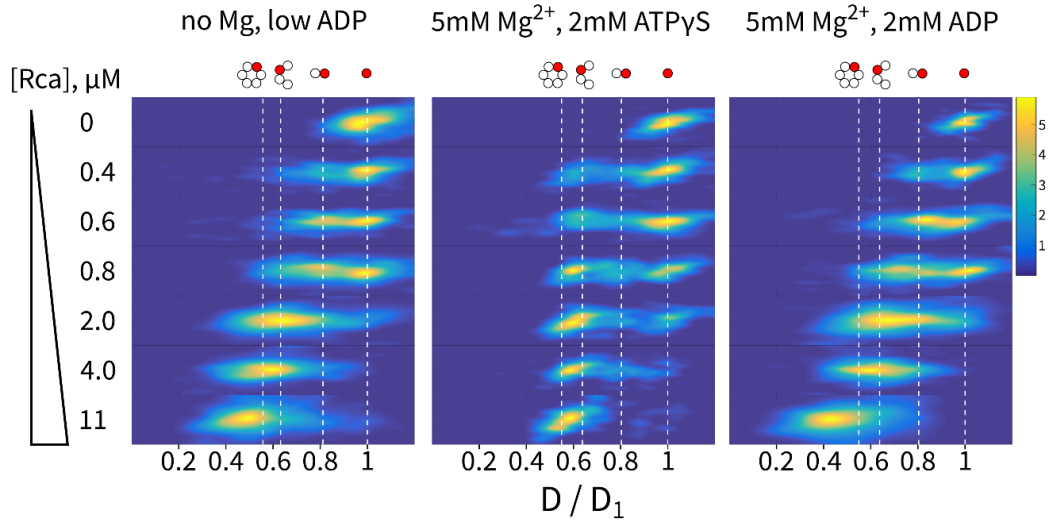


Figure 5.2 Mapping the distribution of Nt-Rca oligomers under different nucleotide binding conditions at varying protein concentrations. Density data are displayed with a normalized color scale (arbitrary unit).

(~14% difference, similar to the limit of our resolution, Figure S6). Between 2 and 11μM,

we observed a gradual sharpening of the tetramer/hexamer population. At 11μM, Nt-Rca seemed to be "locked" into the tetramer/hexamer form, instead of continuing to form higher oligomers, as observed for the low-ADP condition (Figure 2 left). We then conducted measurements in the presence of 5 mM Mg^{2+} and 2 mM ADP, which may mimic the post-hydrolysis state of the enzyme complex. In this case, we observed an oligomerization progression very similar to that of the low-ADP protein. Dimers populate at sub-μM concentrations, tetramers and hexamers exist between 1-4μM and aggregates larger than hexamers form at >10μM.

To further quantify the different oligomerization behavior as a function of nucleotide

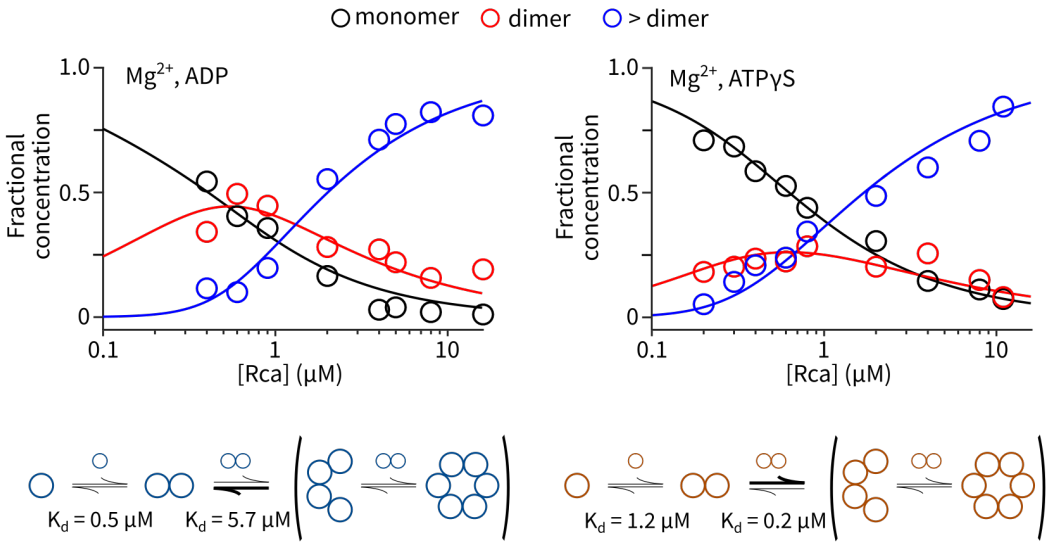


Figure 5.3 Quantifying the oligomerization equilibrium under Mg·ATPγS and Mg·ADP conditions. Solid curves are fits to the assembly model illustrated below.

binding, we first extracted the fractional concentrations of the different oligomers, defined to be $P^n = nC_n^{tot} / \sum_k kC_k^{tot}$ (C_n^{tot} is the total concentration of an n-mer), directly from

fitting the projected D histogram with three Gaussian components (Figure 1A, Figure S3 and Supplementary Tables 1-3). The acquired fractional concentrations (P^n) were then fit to an assembly model defined by the monomer-dimer-tetramer-hexamer pathway using an error minimization procedure (Data analysis section of Materials and Methods). Due to our limited ability to separate tetramers and hexamers, we combined them into one species in the quantification (Figure 3 blue circles). We extracted a monomer-dimer dissociation constant of $K_d^{1-2} = 1.2\,\mu$M with Mg·ATPγS and $K_d^{1-2} = 0.5\,\mu$M with Mg·ADP (residuals of the fit are plotted in Figure S4). On the other hand, the dimer-tetramer/hexamer dissociation constants were an order of magnitude smaller with Mg·ATPγS compared to Mg·ADP ($K_d^{2-4} = 0.2\,\mu$M with Mg·ATPγS and $K_d^{2-4} = 5.7\,\mu$M with Mg·ADP), indicating that ATPγS induces rapid formation of tetramer/hexamers from dimers.

Assembly and disassembly of Nt-Rca are dynamic processes on the 1-10s timescale

Single-molecule diffusometry with the ABEL-oligo method offers the unique opportunity to observe dynamic assembly/disassembly processes that underpin an equilibrium, by measuring the time-dependent changes in hydrodynamic radius of a single protein complex. To observe these processes in real time, we optimized experimental conditions (Materials and Method) to greatly extend trapping duration of single Alexa647-labeled Nt-Rca proteins (from ~1 second to ~5-10 seconds). We indeed observed many molecules showing dynamic transitions in diffusivity with time resolution of ~150ms. Figure 4A shows two examples, both taken with Mg·ADP. In the top panel, the molecule entered the trap as a tetramer (D~30 $\mu m^2/s$ or $0.6D_1$ from 0.2 to 0.8s), disassembled into a monomer (D~50 $\mu m^2/s$ from 0.8 to 1.8s), and then underwent sequential assembly (monomer-dimer-

tetramer/hexamer, from1.8s to the end of the trace). In the bottom panel, the molecule most likely entered as a dimer (D~42 $\mu m^2/s$ or $0.8D_1$) and underwent multiple rounds of

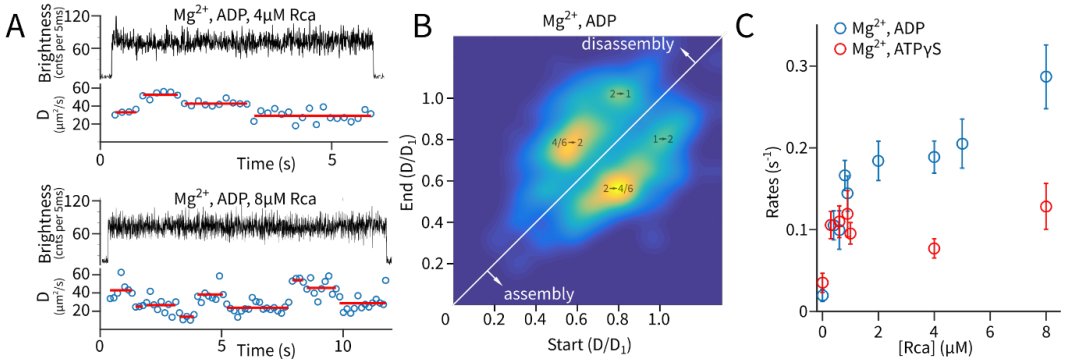


Figure 5.4: Dynamic assembly and disassembly processes observed at the single-molecule level. (A) Two example traces showing the single-molecule diffusion coefficient stepping between discrete levels. (B) Density map of all resolved transitions under the Mg·ADP condition. The first four density maxima are annotated with the interpreted transition pathways. (C) Transition rates versus protein concentration, under different nucleotide binding conditions. Error bars were estimated from the Poisson counting noise of the number of transitions.

assembly and disassembly processes. By pooling all transitions detected under the Mg·ADP condition, we created a transition map by plotting the ending D state against the starting D state (Figure 4B). As illustrated by the transition map, the most encountered transitions are between dimers and tetramers/hexamers. The second most populated transition is between monomers and dimers. These results highlight the dynamic nature of the dimer-tetramer/hexamer pathway and further support the 1-2-4-6 assembly model we used to fit the equilibrium data. Interestingly, dynamic assembly and disassembly processes were also observed with Mg·ATPγS but with much decreased frequency, as quantified by Figure 4C. From the frequencies of the observed transitions, we estimated the rate of subunit exchange to be ~$0.1s^{-1}$ with ATPγS and >$0.2\ s^{-1}$ with ADP. These data suggest that binding of ATPγS locks the protein into a more stable oligomer while ADP binding destabilizes the oligomers. It also suggests that some of the smearing observed in Figure 2,

especially in the Mg·ADP case (for example, the trimer-like density maxima at 0.8μM), could be due to assembly/disassembly dynamics.

Discussion

We have shown that single-molecule diffusometry is a novel and powerful method to resolve protein oligomerization pathways in solution, under equilibrium conditions. It is a single-molecule (here single-complex) method, thus applicable to sub-μM concentrations and dynamically interacting samples, where traditional bulk techniques fail. The method is based on measuring a physical property (the hydrodynamic radius) of individual objects, thus providing a more direct readout compared to other single-molecule fluorescence techniques (e.g. step-wise photobleaching, FRET, etc.). The method only requires fluorescent labeling of a small fraction of the species in equilibrium and can monitor oligomerization reactions at physiological protein concentrations (>1 μM), unlimited by the concentration barrier[40] for conventional single-molecule optical detection. Here, we demonstrated sufficient resolution to resolve monomers, dimers, tetramer/hexamers and higher oligomers. Indeed, the method will be most powerful in the small oligomer regime, where the hydrodynamic sizes of the different species differ by >20%. It should be noted that with a sufficient amount of detected photons, much smaller differences could in principle be resolved[31]. Taken together, we envision the ABEL-oligo method described here to be widely applicable to other oligomerization systems and more generally, to problems that involve biomolecular interactions.

The biological system studied here is the assembly pathway of tobacco Rubisco activase. Our data is consistent with previous measurements on the bulk level (Figure S5), and

reveals rich new details at the single-molecule level. We find that Nt-Rca assembles through a monomer-dimer-tetramer-hexamer pathway, which is consistent with the recent biophysical characterization of cotton Rca using FCS[14,24] and related AAA+ family proteins[41]. We find that ATP binding (mimicked by non-hydrolyzable ATPγS in our experiments) strongly influences the assembly pathway. In particular, ATP binding promotes assembly of higher-order oligomers in the form of tetramers and hexamers. Notably, the extracted equilibrium constant for hexamer formation from dimers ($K_d^{2\text{-}4} \times K_d^{4\text{-}6}$), 0.2 □$M^2$, is identical to that reported for cotton □-Rca in the presence of ATP□S[24]. Quantitative analysis of the equilibrium composition and dynamic assembly/disassembly transitions of Nt Rca reveal the dimer as a critical intermediate that responds differentially to ATPγS and ADP binding. ATPγS binding drastically reduces the dimer-tetramer/hexamer dissociation constant and suppresses the rate of subunit exchange, leading to efficient and stable formation of tetramers/hexamers that dominate the equilibrium at 1-10μM protein concentration.

We hypothesize that ATP binding at the subunit-subunit interface induces a conformational change at the dimer level that "primes" the protein for stable formation of higher-order, functional complexes. Nucleotide binding could serve as a regulatory mechanism that controls the concentration of functional oligomers. We anticipate future experiments on mutants and under different ATP/ADP ratios, together with correlated activity assays, to reveal deeper structural-functional insight into this system.

References

1. Goodsell, D. S. & Olson, A. J. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 105–153 (2000).

2. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.* **6,** 519–529 (2005).

3. Stinson, B. M. *et al.* Nucleotide Binding and Conformational Switching in the Hexameric Ring of a AAA+ Machine. *Cell* **153,** 628–639 (2013).

4. Novák, B. & Tyson, J. J. Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9,** 981–991 (2008).

5. Jiang, Y. *et al.* Sensing cooperativity in ATP hydrolysis for single multisubunit enzymes in solution. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 16962–16967 (2011).

6. Haass, C. & Selkoe, D. J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat. Rev. Mol. Cell Biol.* **8,** 101–12 (2007).

7. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat Med* **10 Suppl,** S10--S17 (2004).

8. Henderson, J. N., Hazra, S., Dunkle, A. M., Salvucci, M. E. & Wachter, R. M. Biophysical characterization of higher plant Rubisco activase. *Biochim. Biophys. Acta - Proteins Proteomics* **1834,** 87–97 (2013).

9. Keown, J. R., Griffin, M. D. W., Mertens, H. D. T. & Pearce, F. G. Small oligomers of ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase are required for biological activity. *J. Biol. Chem.* **288,** 20607–20615 (2013).

10. Keown, J. R. & Pearce, F. G. Characterization of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase activase isoforms reveals hexameric assemblies with increased thermal stability. *Biochem. J.* **464,** 413–423 (2014).

11. Ostanevich, Y. M. *et al.* Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys* **66,** (2003).

12. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8,** 409–427 (2016).

13. Leney, A. C. & Heck, A. J. R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28,** 5–13 (2017).

14. Chakraborty, M. *et al.* Protein Oligomerization Monitored by Fluorescence Fluctuation Spectroscopy: Self-Assembly of Rubisco Activase. *Biophys. J.* **103,** 949–958 (2012).

15. Rajagopalan, S., Huang, F. & Fersht, A. R. Single-molecule characterization of oligomerization kinetics and equilibria of the tumor suppressor p53. *Nucleic Acids Res.* **39,** 2294–2303 (2011).

16. Meseth, U., Wohland, T., Rigler, R. & Vogel, H. Resolution of Fluorescence Correlation Measurements. *Biophys. J.* **76,** 1619–1631 (1999).

17. Orte, A. *et al.* Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. *Proc. Natl. Acad. Sci.* **105,** 14424–14429 (2008).

18. Zijlstra, N., Claessens, M. M. A. E., Blum, C. & Subramaniam, V. Elucidating the Aggregation Number of Dopamine-Induced α-Synuclein Oligomeric Assemblies. *Biophys. J.* **106,** 440–446 (2014).

19. Tosatto, L. *et al.* Single-molecule FRET studies on alpha-synuclein oligomerization of Parkinson's disease genetically related mutants. *Nat. Publ. Gr.* (2015). doi:10.1038/srep16696

20. Portis, A. R. Rubisco activase - Rubisco's catalytic chaperone. *Photosynth. Res.* **75,** 11–27 (2003).

21. Parry, M. A. J. *et al.* Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* **64,** 717–730 (2013).

22. Bhat, J. Y., Thieulin-Pardo, G., Hartl, F. U. & Hayer-Hartl, M. Rubisco Activases: AAA+ Chaperones Adapted to Enzyme Repair. *Front. Mol. Biosci.* **4,** 20 (2017).

23. Wang, Z. Y., Ramage, R. T. & Portis, A. R. Mg 2+ and ATP or adenosine 5'-[y-thio]-triphosphate (ATPyS) enhances intrinsic fluorescence and induces aggregation which increases the activity of spinach Rubisco activase. *Biochim. Biophys. Acta* **1202,** 47–55 (1993).

24. Kuriata, A. M. *et al.* ATP and magnesium promote cotton short-form ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) activase hexamer formation at low micromolar concentrations. *Biochemistry* **53,** 7232–7246 (2014).

25. Stotz, M. *et al.* Structure of green-type Rubisco activase from tobacco. *Nat. Struct. Mol. Biol.* **18,** 1366–1370 (2011).

26. Hasse, D., Larsson, A. M. & Andersson, I. Structure of Arabidopsis thaliana Rubisco activase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71,** 800–808 (2015).

27. Baker, R. T. *et al.* in *Ubiquitin and Protein Degradation, Part A* (ed. Deshaies, R. J.) 540–554 (2005).

28. Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* **13,** 1331–1339 (2004).

29. Peterson-Forbrook, D. S. *et al.* Nucleotide Dependence of Subunit Rearrangements in Short-Form Rubisco Activase from Spinach. *Biochemistry* **56,** 4906–4921 (2017).

30. Wang, Q. & Moerner, W. E. An adaptive anti-Brownian electrokinetic trap with real-time information on single-molecule diffusivity and mobility. *ACS Nano* **5,** 5792–5799 (2011).

31. Wang, Q. & Moerner, W. E. Single-molecule motions enable direct visualization of biomolecular interactions in solution. *Nat. Methods* **11,** 555–558 (2014).

32. Aitken, C. E. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94,** 1826–1835 (2008).

33. Decher, G. Fuzzy Nanoassemblies: Toward Layered Polymeric Multicomposites. *Science (80-.).* **277,** 1232–1237 (1997).

34. Kartalov, E. P., Unger, M. A. & Quake, S. R. Polyelectrolyte surface interface for single-molecule fluorescence studies of DNA polymerase. *Biotechniques* **34,** 505–510 (2003).

35. Watkins, L. P. & Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B* **109,** 617–628 (2005).

36. Botev, Z. I., Grotowski, J. F. & Kroese, D. P. Kernel density estimation via diffusion. *Ann. Stat.* **38,** 2916–2957 (2010).

37. Ortega, A., Amorós, D. & García de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **101,** 892–898 (2011).

38. Tyn, M. T. & Gusek, T. W. Prediction of diffusion coefficients of proteins. *Biotechnol. Bioeng.* **35,** 327–338 (1990).

39. Hazra, S., Henderson, J. N., Liles, K., Hilton, M. T. & Wachter, R. M. Regulation of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) activase: Product

inhibition, cooperativity, and magnesium activation. *J. Biol. Chem.* **290,** 24222–24236 (2015).

40. Holzmeister, P., Acuna, G. P., Grohmann, D. & Tinnefeld, P. Breaking the concentration limit of optical single-molecule detection. *Chem. Soc. Rev.* **43,** 1014–1028 (2014).

41. Kress, W., Mutschler, H. & Weber-Ban, E. Assembly pathway of an AAA+ protein: Tracking ClpA and ClpAP complex formation in real time. *Biochemistry* **46,** 6183–6193 (2007).

CHAPTER 6

CONCLUSIONS / FUTURE DIRECTION

The work presented in the previous chapters of this thesis reflect those projects which have been brought to completion, at least in part. However, in addition to those projects, others have been carried out that have not reached completion and thus require additional experimentation. Below are listed three projects which are currently under active investigation in the Wachter and Levitus labs, all of which were started during my doctoral work.

Assembly of Rca in the Presence of Rubisco:

Until recently, efforts were directed on understanding the self-assembly mechanism of higher plant Rca in the absence of Rubisco. However, because we know that within the chloroplast, Rubisco and Rca are present in relatively high abundance, it may be the case that the self-assembly of Rca determined in the absence of Rubisco may differ from that in the presence of Rubisco. Therefore, given our better understanding of the self-assembly mechanism of Rca, work in the Wachter lab is focusing on incorporation of Rubisco into the FCS work. Due to the collaboration with Dr. Salvucci from the USDA, generous amounts of cotton Rubisco were purified from the leaves and given for use in assembly studies. It was therefore decided that rather than investing time into generating Rubisco from other species, work would begin with cotton Rubisco.

In order to properly lay a foundation for the work with both Rca and Rubisco, it was first necessary to generate an FCS profile for Rubisco, as was done previously for Rca. To this end, work began by first labeling Rubisco with the same AlexaFluor 546 dye used with Rca. Unlike Rca, however, the WT Rubisco sample did not have an engineered

cysteine to specifically target, making the labeling procedure less than ideal. In order to properly determine conditions for labeling, several different conditions were screened, each modifying either pH or molar ratios of dye and protein. It was ultimately determined that by keeping the pH at 7.2, and using a 1 to 1 molar ratio of dye to protein, a sample with a labeling ratio of 0.67 to 1 dye to Rubisco holoenzyme, could be generated and used for FCS. Ideally, one would prefer to have a labeling ratio of 1, however, if this is not possible, under-labeling of the protein is preferred over the alternative. With the sample ready for use, generating this profile consisted of measuring Rubisco at different concentrations, ranging from 1 nM to 300 nM, of labeled Rubisco, and extracting D_{app} values for each particular concentration. The FCS methodology used here was identical to that used with Rca discussed previously. This control experiment was carried out to ensure the sample was aggregating or dissociating on the FCS set-up. The data collected from the FCS experiment was fit using the autocorrelation function discussed previously and resulted in very poor fits. The subsequent D_{app} values are quite different from the calculated diffusional coefficient of 34 $\mu m^2 s^{-1}$, ranging from 75 $\mu m^2 s^{-1}$ to 105 $\mu m^2 s^{-1}$. Because the fit was over fitting in the short time scales and under fitting in the longer time scales (**Fig. 6.1A**), it was speculated that there might be two species present in solution. Therefore, the data was re-analyzed using an autocorrelation function accounting for two diffusing species (equation 1), which resulted in substantially better fits, giving D_{app} values consistent with free Alexafluor 546 (D1 320 $\mu m^2 s^{-1}$) and values consistent with L8S8 Rubisco (D2 32 $\mu m^2 s^{-1}$) (**Fig. 6.1B**).

$$G(\tau) = G1\left(1 + \frac{4D1\tau}{{r_0}^2}\right)^{-1}\left(1 + \frac{4D1\tau}{{z_0}^2}\right)^{-1/2} + G2\left(1 + \frac{4D2\tau}{{r_0}^2}\right)^{-1}\left(1 + \frac{4D2\tau}{{z_0}^2}\right)^{-1/2} \quad \text{Equation 1}$$

Although the fit was improved, the relative abundance of each component was problematic. G1, which corresponds to the relative contribution of a sample with a D1, is

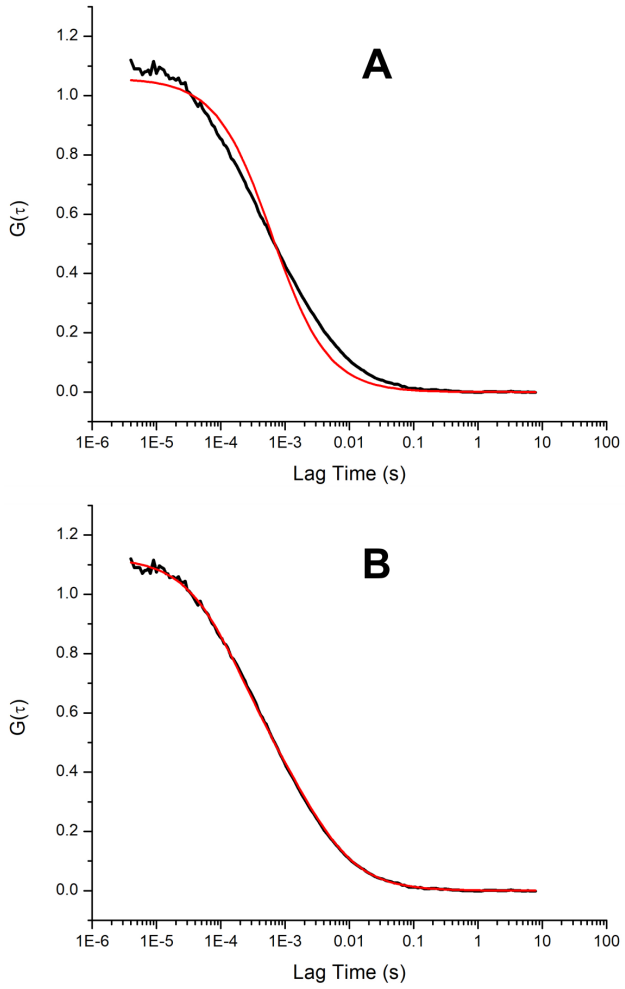


Figure 6.1: Comparison of Rubisco FCS fits using one or two diffusing species. Shown are autocorrelation curves (black curves) generated during FCS experiments on cotton Rubisco at a fixed concentration of 0.1 µM. These curves were fit (red curves) with either an equation for (A) single diffusing species, or (B) two diffusing species.

about 0.51, meaning that more than half of the measured sample is actually free dye, rather than L8S8 Rubisco. This is particularly problematic, given that the labeled Rubisco sample

was run over two G50 Sephadex columns, which for all other labeled Rca samples, removed nearly all of the free dye. Therefore, even though the fit is consistent with the data, it is still indicative of a potential problem.

In an attempt to validate the FCS results, the sample was analyzed by HPLC. The rationale behind the validation was twofold, one, to see whether the dye was attaching to either the large or the small subunit, and two, to determine whether there was truly 50% dye remaining in the sample, as suggested by the fit. Unfortunately, the HPLC work did not help in identifying where the label was actually attaching, as the large and small subunit co-eluted off the column together as one peak. The corresponding peak at the 554 nm signal, which arises from the AlexaFluor 546 dye, was directly in line with the signal detected at 220 nm and 280 nm, which corresponds to Rubisco, confirming the attachment of the dye but not distinguishing between the two subunits. This co-elution was further confirmed by SDS-PAGE analysis. Despite not helping with the first aim, the HPLC work did confirm that there was no free-dye present in the sample used for FCS, suggesting that fitting of the autocorrelation decays with two species was not accurate. Therefore, before work can continue on combining Rca and Rubisco, further characterization of Rubisco alone must be done.

Two practical considerations to keep in mind when continuing this work are changes in solution viscosity and limitations of the perfusion chamber set up. In trying to combine both Rubisco and Rca, the total protein concentration, if too high, may increase the viscosity of the solution and in turn alter the assembly pathway of Rca. A control experiment using either BSA or more glycerol should be considered to ensure that any change in assembly with increased viscosity is indeed from the change in viscosity rather

than just addition of Rubisco. In addition, increased protein concentrations may prove difficult to work with in the current CoverWell perfusion chambers. These chambers tend to form a lot of bubbles when highly concentrated protein is added which may damage the proteins being used.

Characterization of Rubisco Reactivation Using DNA Origami:

Recent developments in the world of DNA manipulation have allowed for the discovery and growth of a field known as DNA Origami [1]. In brief, this field consists of trying to fold either long strands of DNA or circular plasmids into different 3D structures that can be used for different types of applications ranging from crystallography to drug delivery [2]. Currently, work in the Wachter lab is focusing on using this DNA origami technology to position Rubisco and Rca in close proximity on a DNA origami matrix in order to better examine the reactivation mechanism Rca uses in reactivating Rubisco. This work is being carried out in collaboration with the Yan research group in Biodesign at Arizona State University.

The experimental design behind this project involves attaching single stranded DNA (ssDNA) functionalized with a maleimide linker to the engineered cysteine on the C-terminal tail of tobacco Rca. This DNA tag could then be used to attach the Rca at a specific location of the DNA origami matrix which has the complimentary DNA strand. The attachment sites would be placed in six different locations as to form a hexamer when six protomers bind. Due to the flexibility of the DNA linker, each protomer would still be able to toggle to and fro such that it could assemble with the neighboring subunits. Rubisco would also be attached to the DNA origami matrix using the same DNA linking method. It could be positioned at varying distances from Rca as well as in different orientations to

test whether reactivation is more efficient in one particular orientation (**Fig. 6.2**).

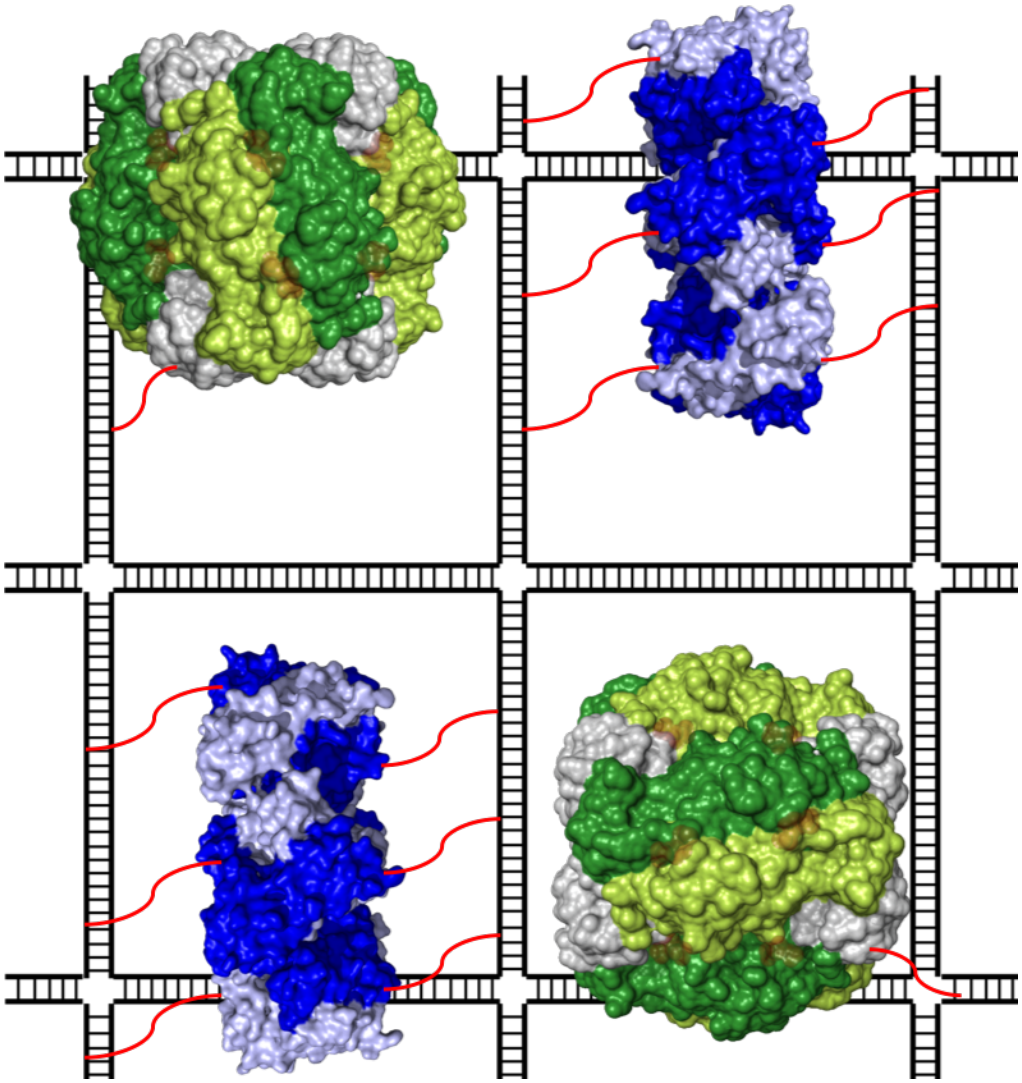


Figure 6.2: Experimental design of Rubisco and Rca attachment to the DNA Origami matrix. Proposed attachment of Rca and Rubisco to a DNA Origami matrix (PDB codes; 3ZW6 and 1RCX). Shown are two possible attachment orientations for Rubisco and Rca, where Rca is interacting with the large subunit of Rubisco (top configuration) or the small subunit (bottom configuration). Other orientations are possible, but not shown.

Progress towards this end has been gradual, but steps in the right direction have been taken. Currently, we have been able to successfully link the ssDNA tag onto tobacco Rca, using a similar method to that used for labeling with a fluorescent tag.

Characterization by RP-HPLC shows that about 70% of Rca used in the reaction gets labeled with the DNA tag. This under-labeling is not as big of an issue given that high excess of protein is used when attaching to the DNA origami matrix. Additionally, all those lacking the tag can not attach and are washed off with buffer. Quantification of total amount labeled is much more difficult than with a fluorescent tag due to DNA and protein both contributing to the 220 nm, 260 nm, and the 280 nm signals. Therefore the comparison was made by loading the same amount of unlabeled protein into the HPLC and using the relative absorbance to extrapolate out the contribution of the DNA tag. This method is far from robust and can certainly use some standardization. Preliminary work on attaching the DNA-Rca complexes to the DNA origami matrix looks promising, however, more work needs to be done on generating a DNA-Rubisco sample to use in conjunction with Rca.

Effect of Mg^{2+} on Rca Assembly:

There are many different factors inside the chloroplast which are thought to regulate the activity and assembly of Rca. Chief among them is fluctuating levels of free Mg^{2+}; levels which increase with increasing incident light and decrease with decreasing light [3]. Magnesium plays a critical role in coordinating the β and γ phosphates of ATP bound in the active site, and has been shown to promote cooperativity in Rca assembly through a secondary binding site [4]. Recent work studying the self-assembly pathway of cotton Rca, found that increasing free Mg^{2+} causes the assembly distribution to favor smaller oligomers, as indicated by having larger D_{app} values. This work was carried out at a fixed protein concentration of 45 μM and in a buffer containing a mixture of ATP and ADP at a ratio of 3:1 (ATP:ADP) [5].

More recent work examining this effect in tobacco Rca found that, like cotton,

increasing free Mg^{2+} also changes the distribution to favor that of smaller oligomers. With

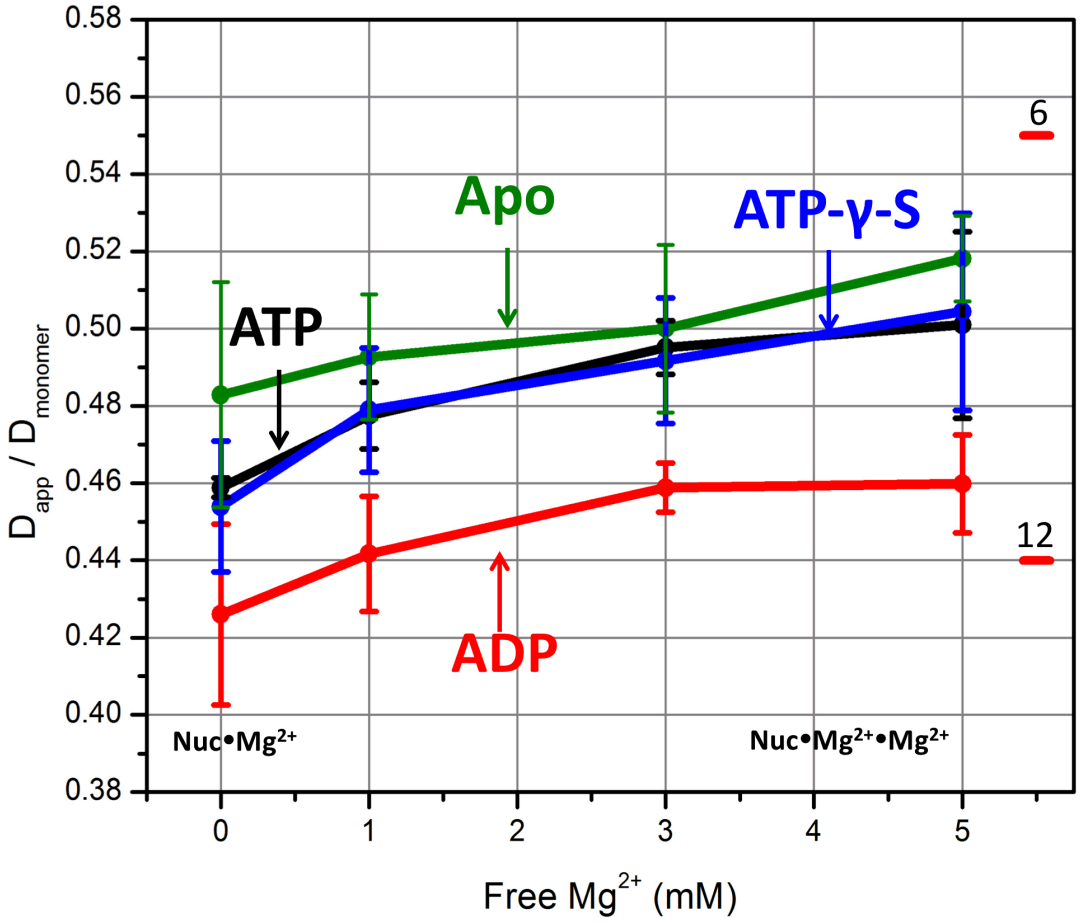


Figure 6.3: Assembly effect of free Mg^{2+} on "Wild-type" Rca in different nucleotides. Changes in apparent diffusion coefficient (D_{app}) with increasing free Mg^{2+} concentration while incubated in different nucleotides, ADP (red data), ATP (black data), ATP-γS (blue data), and Apo (green data). Size markers representing different sized oligomers are shown as red bars on the right hand side and serve as a guide towards understanding the average oligomeric size of the sample. All data collected in at least triplicate.

the tobacco work, a more thorough analysis was carried out given that 4 conditions were examined as opposed to one, like with cotton Rca. The protein concentration was kept constant at 5 μM, however, the sample was incubated in four different conditions: ATP-γS, Active ATP Turnover, ADP, and Apo. Given that incubation in these conditions induce

differences in assembly, all conditions were normalized using a $D_{monomer}$ value determined specifically for that conditions. Despite this variability in assembly, it is clear that all

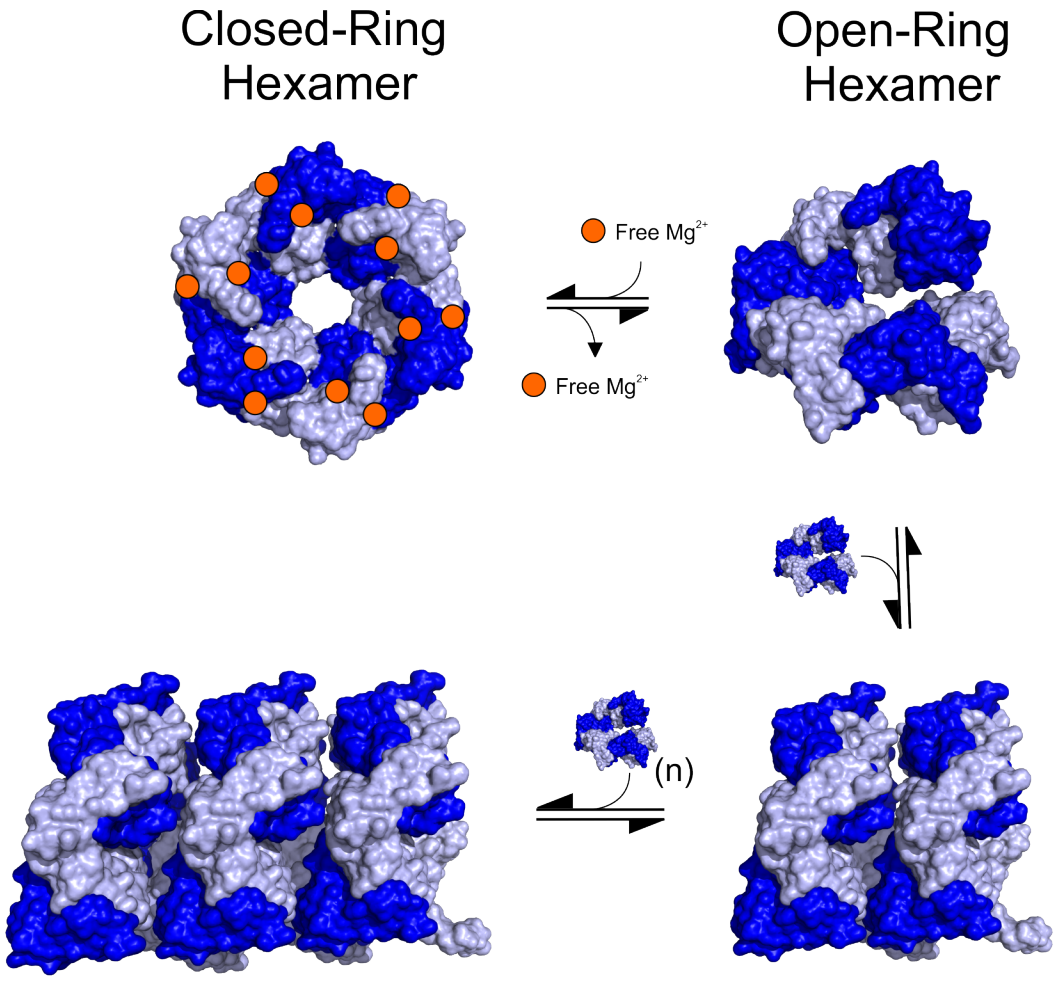


Figure 6.4: Modeling effect of changing free Mg^{2+} levels on assembly. Proposed model showing the changes in assembly induced by changing levels of free Mg^{2+} levels (PDB codes; 3T15 and 3ZW6). High levels of free Mg^{2+} favors a larger population of closed-ring oligomers while decreasing levels favor more open-ring oligomers, allowing for more rapid assembly towards larger oligomers.

conditions show an increase in the $D_{app}/D_{monomer}$ ratio with increasing free Mg^{2+} suggesting that like cotton, the average size of the oligomers are decreasing (**Fig. 6.3**). This finding lead to the development of a model where increasing free Mg^{2+} favors the formation of more closed-ring complexes while decreasing free Mg^{2+} leads to formation of open-ring

complexes thus making assembly to higher ordered oligomers that much more likely (**Fig. 6.4**). It is worth noting that this work focused exclusively on the Mg^{2+} effect on assembly, but work published recently suggests that Manganese (Mn^{2+}) also effects the behavior of spinach Rca in terms of subunit exchange dynamics, which may in turn effect the self-assembly pathway [6]. Therefore further work must be carried out to better establish the foundation for understanding the role all divalents play in regulating green-type Rcas. Additionally, this divalent effect needs to be understood as a function of concentration, not just at one in particular.

Conclusion:

A great deal of work has been done thus far to characterize the self-assembly pathway of higher plant Rca, however, this story is far from compete. Based on the work presented in this thesis, we can say that our understanding of the role nucleotides and divalents play in regulating self-assembly is greater, yet must affirm that there is still much more to learn. I have argued that establishing this foundation for Rca self-assembly is vital, if we hope to stay in step with the ever-increasing demand for food by an ever-increasing population, due primarily to Rca's critical interaction with Rubisco. Therefore my hope in writing this thesis is that those pioneers in the activase field will utilize this work and re-engineer Rubisco and Rca into more efficient proteins in order to meet the crop demands of the future. Not one thing worth pursuing is ever easy, a saying that was never more true than when it is used to describe work being done on activases.

References

1. Yan, H. *et al.* 2003. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science* 301, 1882-1884

2. Jiang, Q. *et al.* (2012) DNA Origami as a Carrier for Circumvention of Drug Resistance. *J. Am. Chem. Soc.* **134,** 13396-13403

3. Carmo-Silva, A. E. and Salvucci, M. E. 2013. The Regulatory Properties of Rubisco Activase Differ Among Species and Affect Photosynthetic Induction During Light Transitions. *Plant Physiol.* **161,** 1645-1655

4. Hazra, S., *et al*. 2015. Regulation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase: Product Inhibition, Cooperativity, and Magnesium Activation. *J. Biol. Chem*. **290,** 24222-24236

5. Kuriata, A. M. *et al*. 2014. ATP and Magnesium Promote Cotton Short-Form Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Hexamer Formation at Low Micromolar Concentrations. *Biochemistry* 53, 7232-7246.

6. Peterson-Forbrook, D. S., *et al*. 2017. Nucleotide Dependence of Subunit Rearrangements in Short-Form Rubisco Activase from Spinach. *Biochemistry* 56, 4906-4921

APPENDIX A

COPYRIGHTS TO FIGURES USED IN THE INTRODUCTION CHAPTER

Copyrights for Figure 1.1

Copyright Clearance Center

Note: Copyright.com supplies permissions but not the copyrighted content itself.

1 PAYMENT 2 REVIEW **3 CONFIRMATION**

Step 3: Order Confirmation

Thank you for your order! A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us 24 hrs/day, M-F at +1.855.239.3415 Toll Free, or write to us at info@copyright.com. This is not an invoice.

Confirmation Number: 11724327
Order Date: 06/15/2018

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

Andrew Serban
ajserban@asu.edu
+1 (602) 791-8938
Payment Method: n/a

Order Details

Annual review of plant biology

Order detail ID:	71246950
Order License Id:	4370540530524
ISSN:	1545-2123
Publication Type:	e-Journal
Volume:	
Issue:	
Start page:	
Publisher:	ANNUAL REVIEWS
Author/Editor:	Annual Reviews, Inc

Permission Status: Granted
Permission type: Republish or display content
Type of use: Thesis/Dissertation

Requestor type	Academic institution
Format	Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/figures	1
The requesting person/organization	Andrew Serban
Title or numeric reference of the portion(s)	Chapter 1, Figure 1.1

Copyrights for Figure 1.3

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jun 15, 2018

This Agreement between Mr. Andrew Serban ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4370540812297
License date	Jun 15, 2018
Licensed content publisher	Oxford University Press
Licensed content publication	Journal of Experimental Botany
Licensed content title	Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships
Licensed content author	Tabita, F. Robert; Satagopan, Sriram
Licensed content date	Feb 16, 2008
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Publisher of your work	n/a
Expected publication date	Jun 2018
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Instructor name	n/a
Institution name	n/a
Expected presentation date	Jun 2018
Portions	Chapter 1, Figure 1.3
Requestor Location	Mr. Andrew Serban 1710 N 114th Ave AVONDALE, AZ 85392 United States Attn: Mr. Andrew Serban
Publisher Tax ID	GB125506730

Copyrights for Figure 1.4

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jun 15, 2018

This Agreement between Mr. Andrew Serban ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4370540969866
License date	Jun 15, 2018
Licensed content publisher	Oxford University Press
Licensed content publication	Journal of Experimental Botany
Licensed content title	Catalysis and regulation in Rubisco
Licensed content author	Andersson, Inger
Licensed content date	Apr 15, 2008
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Publisher of your work	n/a
Expected publication date	Jun 2018
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Instructor name	n/a
Institution name	n/a
Expected presentation date	Jun 2018
Portions	Chapter 1, Figure 1.4
Requestor Location	Mr. Andrew Serban 1710 N 114th Ave AVONDALE, AZ 85392 United States Attn: Mr. Andrew Serban
Publisher Tax ID	GB125506730
Billing Type	Invoice

Copyrights for Figure 1.6

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jun 15, 2018

This Agreement between Mr. Andrew Serban ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4370550273906
License date	Jun 15, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Structural & Molecular Biology
Licensed Content Title	Structure of green-type Rubisco activase from tobacco
Licensed Content Author	Mathias Stotz, Oliver Mueller-Cajar, Susanne Ciniawsky, Petra Wendler, F Ulrich Hartl et al.
Licensed Content Date	Nov 6, 2011
Licensed Content Volume	18
Licensed Content Issue	12
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Instructor name	n/a
Institution name	n/a
Expected presentation date	Jun 2018
Portions	Chapter 1, Figure 1.6
Requestor Location	Mr. Andrew Serban 1710 N 114th Ave

Copyrights for Figure 1.7

OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Jun 15, 2018

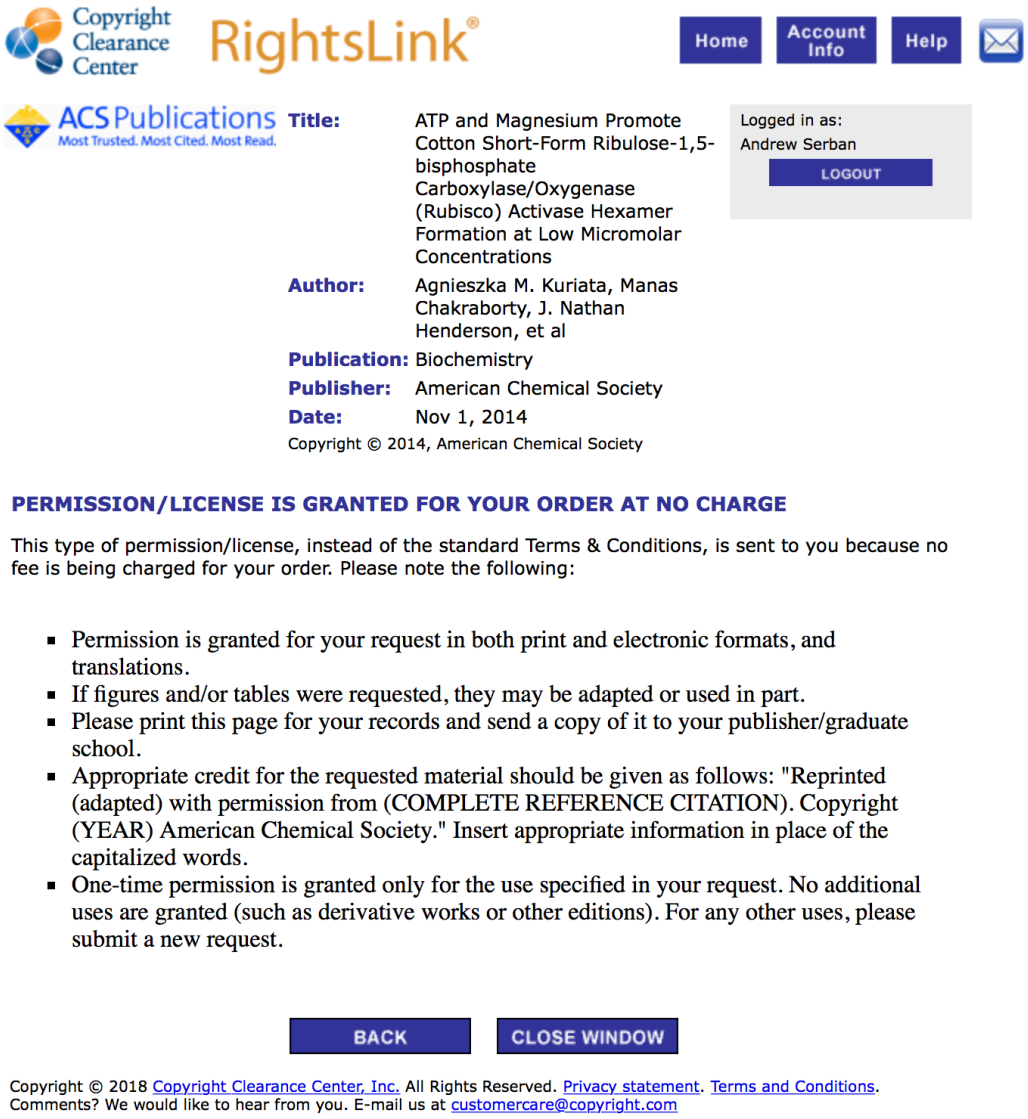
This Agreement between Mr. Andrew Serban ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4370550401832
License date	Jun 15, 2018
Licensed content publisher	Oxford University Press
Licensed content publication	Journal of Experimental Botany
Licensed content title	Regulation of Rubisco activase and its interaction with Rubisco
Licensed content author	Portis, Archie R.; Li, Cishan
Licensed content date	Nov 29, 2007
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Publisher of your work	n/a
Expected publication date	Jun 2018
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Instructor name	n/a
Institution name	n/a
Expected presentation date	Jun 2018
Portions	Chapter 1, Figure 1.7
Requestor Location	Mr. Andrew Serban 1710 N 114th Ave AVONDALE, AZ 85392 United States Attn: Mr. Andrew Serban
Publisher Tax ID	GB125506730
Billing Type	Invoice

APPENDIX B

COPYRIGHTS TO THE KURIATA *ET AL* PUBLICATION

Copyrights for using the Kuriata *et al.* publication



Copyright Clearance Center RightsLink®

Home | Account Info | Help

ACS Publications
Most Trusted. Most Cited. Most Read.

Logged in as:
Andrew Serban
LOGOUT

Title: ATP and Magnesium Promote Cotton Short-Form Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Hexamer Formation at Low Micromolar Concentrations

Author: Agnieszka M. Kuriata, Manas Chakraborty, J. Nathan Henderson, et al

Publication: Biochemistry

Publisher: American Chemical Society

Date: Nov 1, 2014

Copyright © 2014, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK | CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

APPENDIX C

COPYRIGHTS TO THE WANG *ET AL* PUBLICATION

Copyrights for using the Wang *et al.* publication

AIP PUBLISHING LICENSE
TERMS AND CONDITIONS

Jun 02, 2018

This Agreement between Mr. Andrew Serban ("You") and AIP Publishing ("AIP Publishing") consists of your license details and the terms and conditions provided by AIP Publishing and Copyright Clearance Center.

License Number	4352840204057
License date	May 19, 2018
Licensed Content Publisher	AIP Publishing
Licensed Content Publication	Journal of Chemical Physics
Licensed Content Title	Single-molecule diffusometry reveals the nucleotide-dependent oligomerization pathways of Nicotiana tabacum Rubisco activase
Licensed Content Author	Quan Wang, Andrew J. Serban, Rebekka M. Wachter, et al
Licensed Content Date	Mar 28, 2018
Licensed Content Volume	148
Licensed Content Issue	12
Type of Use	Thesis/Dissertation
Requestor type	Author (original article)
Format	Print and electronic
Portion	Excerpt (> 800 words)
Will you be translating?	No
Title of your thesis / dissertation	Understanding the Self-assembly Pathway of Higher Plant Rubisco Activase
Expected completion date	Jun 2018
Estimated size (number of pages)	130
Requestor Location	Mr. Andrew Serban 1710 N 114th Ave AVONDALE, AZ 85392 United States Attn: Mr. Andrew Serban
Billing Type	Invoice
Billing Address	Mr. Andrew Serban 1710 N 114th Ave AVONDALE, AZ 85392

BIOGRAPHICAL SKETCH

Andrew J. Serban was born and raised in Phoenix, Arizona. He went to Westview high school and then attended college at Arizona State University, majoring in Biochemsitry. He got married to his lovely wife Cristina in November 2012 and graduated *Summa cum Laude* shortly thereafter in December of 2012. After graduation, he began working as a laboratory technician at a company called Arizona Nutritional Supplements. After a brief period of time, he went back to Arizona State University to begin working on his PhD as part of a joint project between the Wachter and Levitus researcher groups. During the course of his graduate career, he and his wife had their first child, Jason, and are currently expecting their second child. He finished his PhD in August 2018. He has accepted a job offer for a Director of Operations position at a local Aerospace company named WG Henschen and is excited to begin this new role.